

Article

Fire and Smoke Detection Using Fine-Tuned YOLOv8 and YOLOv7 Deep Models

Mohamed Chetoui  and Moulay A. Akhloufi * 

Perception, Robotics, and Intelligent Machines Laboratory (PRIME), Department of Computer Science, Université de Moncton, Moncton, NB E1A 3E9, Canada; emc7409@umoncton.ca

* Correspondence: moulay.akhloufi@umoncton.ca

Abstract: Viewed as a significant natural disaster, wildfires present a serious threat to human communities, wildlife, and forest ecosystems. The frequency of wildfire occurrences has increased recently, with the impacts of global warming and human interaction with the environment playing pivotal roles. Addressing this challenge necessitates the ability of firefighters to promptly identify fires based on early signs of smoke, allowing them to intervene and prevent further spread. In this work, we adapted and optimized recent deep learning object detection, namely YOLOv8 and YOLOv7 models, for the detection of smoke and fire. Our approach involved utilizing a dataset comprising over 11,000 images for smoke and fires. The YOLOv8 models successfully identified fire and smoke, achieving a mAP:50 of 92.6%, a precision score of 83.7%, and a recall of 95.2%. The results were compared with a YOLOv6 with large model, Faster-RCNN, and DETECTION TRansformer. The obtained scores confirm the potential of the proposed models for wide application and promotion in the fire safety industry.

Keywords: fire detection; wildfires; deep learning; object detection



Citation: Chetoui, M.; Akhloufi, M.A. Fire and Smoke Detection Using Fine-Tuned YOLOv8 and YOLOv7 Deep Models. *Fire* **2024**, *7*, 135. <https://doi.org/10.3390/fire7040135>

Academic Editors: Alistair M. S. Smith, Houda Harkat, José M. P. do Nascimento, Saad Dosse Bennani and Hasmath Farhana Thariq Ahmed

Received: 11 February 2024

Revised: 3 April 2024

Accepted: 8 April 2024

Published: 12 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As an essential component of the Earth's ecology, forests support the growth of human society and the economy while also aiding in the preservation of the planet's natural equilibrium [1]. However, the frequency of forest fires in recent years has put the safety of forested areas at risk. Every year, these fires generate enormous losses for animals and plants since they destroy millions of hectares of forest and damage the biological ecosystem [2]. Australia saw devastating wildfires in 2020 that destroyed thousands of structures, scorched at least 19 million hectares, and killed 34 humans in addition to over a million animals [3]. Typically, forest fires are intense, devastating, and challenging to fight. It is hard to put out a forest fire quickly because once it starts, it can spread in a wide area. If forest fires are detected in real time and areas of fire and smoke are identified rapidly, firefighters will be able to take the necessary measures and control the spread of fires. Thus, it is essential for efficient fire management to know how to detect smoke and fires accurately.

In the field of forest fire monitoring, standard methods require the utilization of manual observation techniques, such as ground-based inspections and survey towers. The process of manual observation is sensitive to various external factors, including logistical limitations, problems with communication, and weather conditions, which can result in inefficiencies. Observation towers have limitations in terms of monitoring, such as limited coverage, areas without surveillance, and high maintenance costs. Although satellite-based monitoring of forest fires has extensive coverage, it suffers several limitations. These include insufficient spatial resolution of satellite imagery, dependency on orbital cycles, vulnerability to interference from weather and cloud cover, and a limited number of satellites.

Contact fire detectors, such as those based on chemical sensors [4] and smoke detectors [5], are the primary sensor detecting tools used in indoor fires and smoke detection. However,

this approach has limitations when it comes to detecting large-scale fires because uncertainties in the surrounding environment, tree shading, sensor orientations, and sensor distances can all have an impact on the detection effect. When it comes to giving early warning of fires, smoke and fire detection systems are not very effective. A large number of video surveillance systems have been installed and used in forest fire early warning systems due to the rapid development of computer vision technology.

The advantages of video surveillance fire detection technology over traditional fire detection technology include contactless operation, an extensive detection range, cheap maintenance costs, rapid response times, and good detection performance. Extracting the visual characteristics of smoke and fire, such as color [6], texture [7], motion [8], the background contrast [9], and the combination of several visual aspects, is the main objective of this approach. Nevertheless, there are numerous issues with the standard machine learning-based fire and smoke detection methods, including low detection accuracy, weak generalization, small pixel characteristics, and complicated forest fire image backgrounds. Since 2012, deep learning techniques for computer vision target detection have advanced rapidly. This is largely due to the development of computational power, particularly with the use of Graphics Processing Units (GPUs), and the availability of public datasets. These developments have resulted in significant improvements in accuracy and efficiency as well as lower computational costs. As a result, deep learning is now widely used in the field of smoke and fire detection. These algorithms are basically divided into two categories: The first is the two-stage algorithm, which has a two-stage model structure. Prior to performing the task categorization and positioning, the ROI (region of interest) candidate region is established. The traditional algorithms are Fast-R-CNN [10] and Faster-R-CNN [11]. The other is the one-stage approach, which directly predicts the target's category and position information via a regression-based target detection network. whose traditional techniques are the SSD (single-shot multi-box detector) algorithm [12] and the YOLO (you only look once) series [13].

An enhanced technique of convolutional neural network (CNN) for smoke and fire detection has been suggested by researchers [14–16]. In order to enhance neural networks, researchers have recently used Transformer Backbone networks [17]; common examples of these are ViT (vision transformer) [18], Swin [19], and PVT (pyramid vision transformer) [20]. An improved method uses Transformer code-based feature fusion for target detection after the pretraining weights from the large-scale image classification data are used as the detector Backbone network's initial weights [21,22]. Although the Transformer model is hard to train and has high computation costs, it performs well. The previously mentioned deep learning-based smoke and fire detection technologies have shown good results; nevertheless, the larger the model, the more convolutional layers in the network structure, and the greater the detection algorithm's performance. In reality, this leads to a model that is undesirable for deployment on devices with limited resources since it has a high number of weighting parameters and a low detection efficiency. Thus, in order to perform smoke and flame detection, lightweight networks are required.

Several studies have been presented recently to detect fire and smoke [23–25]. Talaat et al. [26] proposed an improved fire detection approach for smart cities based on the YOLOv8 object detector algorithm; their approach is called the smart fire detection system (SFDS). The four layers that the authors worked on are the application layer (i), fog layer (ii), cloud layer (iii), and Internet of Things layer (iv). In order to collect and interpret data in real time, the suggested algorithm combines the fog, cloud computing, and the IoT layers. This allows for faster responses and reduces the possibility of damage to individuals or property. Their algorithm achieved in terms of both precision and recall a score of 97.1% for all classes. Al-Smadi et al. [27] provided a framework that reduces the sensitivity of different YOLO detection models. The authors compared the detection performance and speed of different versions of YOLO models such as YOLOv3, YOLOv5, and YOLOv7 with prior ones such as Fast R-CNN and Faster R-CNN. Furthermore, to verify the detection model's accuracy in identifying smoke targets, researchers employed a collected dataset

that describes three different detection areas: close, medium, and long distance. Their model gave a mean average precision at an intersection over union (IoU) threshold of 0.5 (mAP:50) score of 96.8% using YOLOv5x for detecting forest smoke. And YOLOv7 outperformed YOLOv3 with a mAP:50 of 95% vs. 94.8%.

The improved version of YOLOv5, which used the K-means++ algorithm for dynamic anchor learning was developed by Wang et al. [28]. The proposed method aimed to improve the performance and speed of fire detection in order to reduce fire damage. In addition, three different YOLOv5 models—YOLOv5 small, YOLOv5 medium, and YOLOv5 large—were subjected to loss functions, such as CIoU and GIoU. A synthetic approach was trained with a synthetic dataset of 4815 images in order to increase the total number of images to 20,000. The findings show that the mean average accuracy of the modified model was 4.4% higher than that of the original YOLOv5. Furthermore, authors found that YOLOv5 performs better when using the CIoU loss function, with a recall of 78% and a mAP:50 of 87%.

A new method for recognizing fires in aerial photos was reported by Mukhiddinov et al. [29]. A total of 6000 images showing smoke and forest fires were collected from Google and Kaggle. Additionally, the developed method aimed to discover the anchor squares by using the K-means++ algorithm with the YOLOv5 model. Preprocessing, including flipping and rotation, were developed to increase detection accuracy by decreasing the sensitivity of detection models in subsequent detection operations. The authors found a mAP:50 score of 73.6%; the criteria used to evaluate application showed that the provided approach outperformed a number of competing approaches, including SSD and upleNet.

Yazdi et al. [30] focused on the problem of insufficient and poor-quality data for smoke and fire detection in recently published studies. The lack of metadata that is appropriate for the development and use of robust detectors algorithms is one of the most important problems. As a result, a unique strategy was used to establish NEMO (Nevada Smoke Detection Benchmark), a first-of-its-kind data repository consisting of a collection of aerial images collected to identify forest fires from detection stations. NEMO offers datasets containing 7023 fire detection images captured with many cameras at different locations and times. Faster R-CNN and RetinaNet are two of the detection models that were used to assess the data. Within 5 minutes, the findings showed an average detection accuracy of 42.3% and a detection rate of 98.4%.

A unique approach for employing ensemble learning to identify forest fires was introduced by Xu et al. [15]. YOLOv5 and EfficientDet were used as the main learners in the first layer of the proposed technique, followed by EfficientNet, which is in charge of the detection and classification of forest fires. The model trained on 10,581 images from popular datasets like VisiFire and FD-dataset. With a mAP of 79.7% at the intersection over the union of 50% (IoU:50), the results showed an improvement in fire detection accuracy when compared to other models, such as the YOLOv4.

In the study of Abdusalomov et al. [31], a novel method based on the improved YOLOv3 model was proposed for detecting fires both during the day and at night in the shortest amount of time and with the largest possible detection area. The study additionally highlights the lack of reliable data for fire identification. As a result, in addition to compiling a collection of images produced from video clips, a data collection of 9200 photographs was assembled and constructed using Google repositories and remotely accessible resources. Additionally, new copies of the most recent data were created, and the size of the dataset was increased through the use of data-augmentation techniques like image rotation. In this study, the authors focused on the real-time detection of fire. The experimental study had an average accuracy of 98.9%.

In this article, we propose the development of the recent YOLO architecture, which is YOLOv8 for fire and smoke detection in different locations. We fine-tune the models to achieve a better precision, recall, and mean average precision at the intersection over union 50 and 50-95 scores on a large dataset with more than 11,000 images for fire and smoke. We compare the models with previous version of yolo (YOLOv7 and YOLOv6), Faster-RCNN,

and DETection TRansformer (DETR) [32] in terms of precision, recall, mAP:50, and mAP:50-95 scores. Finally, we provide a qualitative analysis of the YOLOv8 in the detection of fire and smoke on different conditions.

The main contributions of this paper are as follows:

1. Presenting a fine-tuned YOLOv8 for smoke and fire detection in various locations.
2. Enhancing precision: The suggested method has the potential to enhance the precision of fire and smoke detection in forests, cities, and other locations when compared to traditional methods. A possibility to achieve this is by using the features of advanced deep learning algorithms like YOLOv8. These algorithms can be trained to recognize and detect specific characteristics of fire and smoke that can be challenging to identify using traditional image-processing techniques.
3. Real time: The YOLOv8 algorithm is recognized for its efficiency and accuracy in real-time object detection. The proposed method is highly suitable for fire and smoke detection applications, where the fast and timely detection of fires is crucial.
4. Large dataset: Instead of using a limited number of images for fire and smoke, this study uses a large dataset that includes fire, smoke, and normal scenes. The dataset contains real-world images collected from multiple sources and includes a variety of fire and smoke scenarios, including both indoor and outdoor fires, varied in size from small to large. A deep CNN extracts important features from the large dataset in order to generate accurate predictions and avoid the problem of overfitting.

2. The Network Structure of YOLOv8

In the field of computer vision, the YOLO model achieved significant interest. Researchers have made improvements and integrated new components into the architecture, resulting in numerous traditional models. YOLOv8, which was released by Ultralytics on 2023, provides an important progression in this development. In comparison to previous models, such as YOLOv5 and YOLOv7, YOLOv8 represents the state of the art for its enhanced detection precision. The network architecture of YOLOv8 consists of three primary components, namely, the neck, backbone, and head.

The CSPDarknet53 model as described in reference [33], is used as the core network in YOLOv8. This model provides five distinct scale features, known as B1–B5, by performing five consecutive downsampling stages. The Cross Stage Partial (CSP) module is changed with the C2f module in the initial architecture of the backbone. The implementation of the C2f module enhances the network architecture by optimizing the shortest and longest gradient paths, resulting in improved training. The CBS module, which means Convolution Batch Normalization Silu, is a composite module that initially appeared in the YOLOv5 architecture for object detection tasks. This module combines three key elements, specifically, convolutional layers are used for the task of extracting features from the input data. Convolutional techniques are performed by these layers to extract basic patterns and features in the data. Batch normalization is a technique proposed to standardize the activations of different layers within a neural network. The decreased number of internal covariate shifts improves the acceleration of the training process. The Silu module (Sigmoid Linear Unit), also referred to as the Swish activation function, is a variant of an activation function which includes non-linear connections within the neural network. The integration of these elements in the CBS module enhances the neural network's performance for the extraction of complex features from the images. In various applications, the integration of this module improves the accuracy and efficacy of object detection. The SPPF layer is used in the last part of the backbone network to generate output of similar size by combining the input feature maps. By using three consecutive maximum pooling layers, SPPF reduces latency and improves computational efficiency when compared to the spatial pyramid pooling structure [34]. The neck component of YOLOv8 contains a PAN-FPN architecture inspired from PANet [35]. YOLOv8 introduces a modification to the neck module of YOLOv5 and YOLOv7 by removing the convolution operation after up-sampling in the PAN structure. This modification keeps the original performance of the model

while achieving an improved configuration. The PAN structure and FPN structure of the YOLOv8 model are characterized by distinct feature scales, which are represented as P4–5 and N4–N5, respectively. Traditional FPN provides deep semantic information using a top–down approach. However, although FPN enhances the integration of semantic information between B4–P4 and B3–P3, it may result in the loss of object localization information. To address this issue, PAN–FPN combines PAN with FPN. The integration of PAN enhances the acquisition of location data by merging P4–N4 and P5–N5, achieving an improvement in the top–down process. This strategy provides a complete network framework that integrates both top–down and bottom–up elements. By combining surface-level positional data with deep semantic details, feature fusion enhances the range and complexity of features. The decoupled head architecture is used by YOLOv8. The structure of the architecture includes different parts related to object classification and bounding box regression prediction. Following that, customized loss functions are used for each individual task. The prediction of bounding box regression using the CIoU [36] and distribution focal loss [37] is a specific task. The binary cross-entropy loss is applied in the classification task. This architecture helps to improve the accuracy of detection and the convergence of the model. Another interesting characteristic of YOLOv8 is its anchor-free detection model, which improves the process of distinguishing between positive and negative samples. Furthermore, the model integrates the Task-Aligned Assigner [38] to dynamically allocate samples in order to improve the detection accuracy and enhance the robustness of the model.

3. Experimental Results

3.1. Dataset

In our study, we used the dataset proposed by [39]. It consists of 11,667 RGB images, and they were split into training, validation, and test sets at random, with 8494, 2114, and 1059 images in each set. According to the authors of the dataset, the two primary sources of the datasets were (1) a web-based collection of videos and photographs, and (2) an independently gathered dataset of tree fires. The final dataset was created by first processing the web-captured and self-collected videos by frame extraction (30 frames apart). The video extracted by frame was then added to the image dataset, which was then labeled using the LabelImg tool [40]. In the image dataset, the scene mostly consists of noon, evening, and night forest fires. The experimental dataset’s cases are classified as “smoke” or “fire”, with “smoke” constituting a comparatively small portion of the dataset. Figure 1 presents some fire and smoke images on the dataset with their labels.



Figure 1. Cont.



Figure 1. Example of images of smoke and fire on the dataset [39].

3.2. Metrics and Hyper-Parameters

Model training and model testing of this paper's experiments were conducted on NVidia V100SXM2 (16G memory) and the CPU Intel Gold 6148 Skylake 2.4 GHz.

In our study, each image has a size of 640×640 and no data augmentation was applied. For each model and for every experimental learning rate, warm-up training was used. In the early training stage's 0 to 3 epochs, the particular technique was to rapidly increase the learning rate from 0 to 0.01. A high learning rate must be achieved early on in the training phase to prevent the oscillation problem in the model, which can make convergence challenging. The weight of the model was updated and optimized via the use of stochastic gradient descent (SGD). The batch size was set to 128; the initial learning rate was set to 0.01; the momentum parameter was set to 0.937; the weight attenuation factor was set to 0.0005; and there were 300 training epochs.

Precision and recall are in our detection tasks. Precision is the percentage of the forecast targets that are correct to all of the predicted targets. Recall shows the percentage of the predicted to all of the forecast targets that correspond to the correct target. The following are the calculating formulas:

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

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

where the numbers for true positive, false positive, and false negative are, respectively, TP , FP , and FN . This experiment uses mAP, a measuring indicator with a value range of 0 to 1, that combines precision and recall.

To evaluate the model's indicators of performance, including mAP, recall, and precision, we set $IoU = 0.5$. The model's weight and speed are evaluated as well using the parameter size and the inference time for each image.

3.3. Results

After training and testing of YOLOv8 and YOLOv7 models, we obtained very interesting scores in terms of mAP:50, precision, and recall on the test dataset. YOLOv8x is the largest model and obtained the highest score with a precision of 0.954, a recall of 0.848 and a mAP:50 of 0.926, followed by YOLOv8l, which obtained a precision score of 0.949, a recall of 0.837, and a mAP:50 score of 0.901. Interesting scores were also obtained by YOLOv8m and YOLOv8s with a mAP:50 of 0.895 and 0.891, respectively. We compared the results with YOLOv7 models. The models also gave a good score in terms of mAP:50, precision, and recall. As we can see, the YOLOv7-E6 model obtained scores very close to the YOLOv8m in terms of precision, mAP:50, and mAP:50-95. Table 1 shows the remaining

results provided by YOLOv8 and YOLOv7. The YOLOv8n model has the lowest score for precision, recall, mAP:0.5, and mAP:0.5-0.95 metrics for the YOLOv8 versions. YOLOv6l provided a mAP:0.5 score close to YOLOv7. Faster-RCNN models and DETR obtained the lowest scores compared to all yolo versions. Figures 2–4 show the precision, recall, and mAP:50 curves on the training process; as we can see, the precision for the YOLOv8x reaches 0.995 for precision.

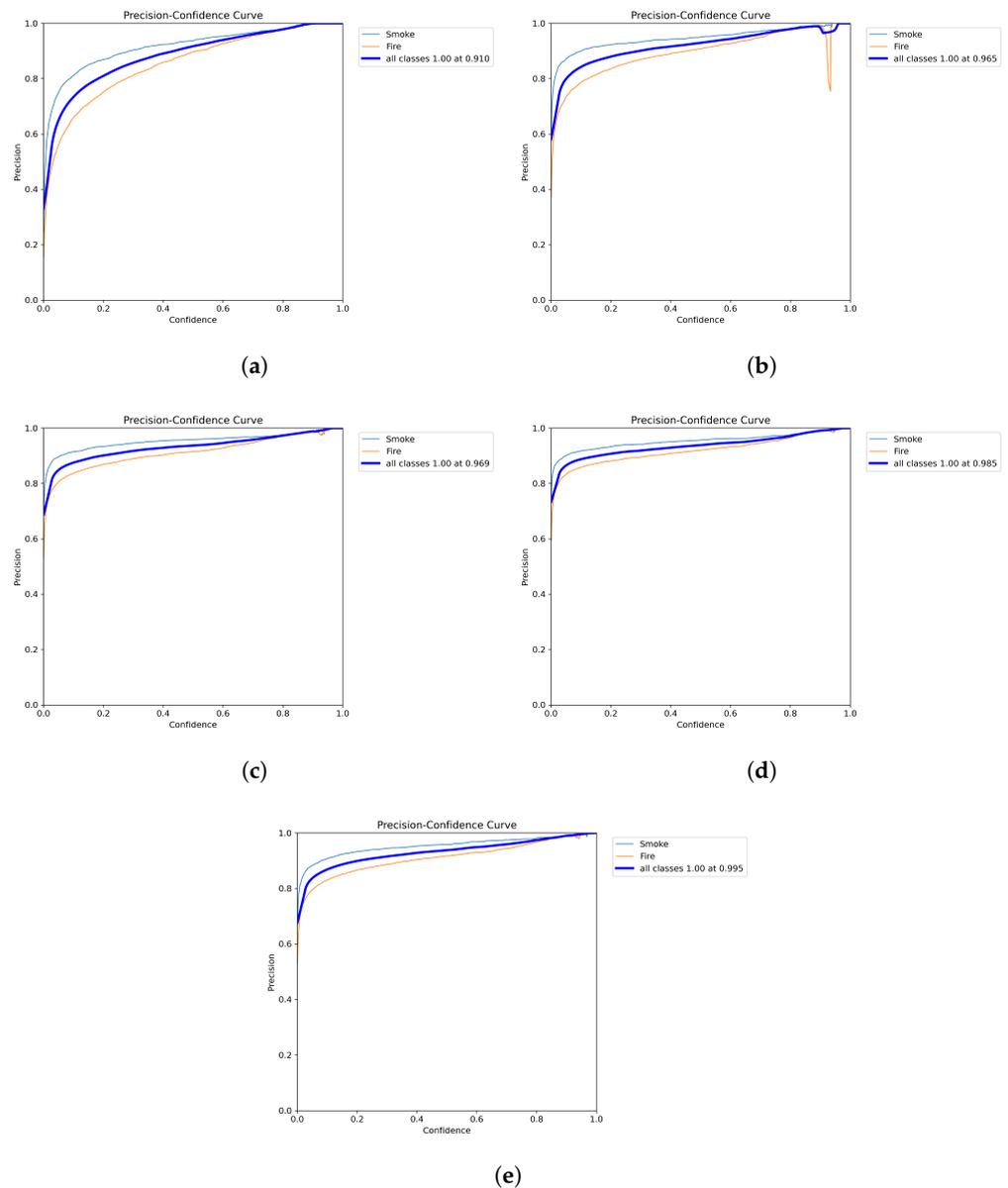


Figure 2. Precision curves of YOLOv8 models: (a) YOLOv8n, (b) YOLOv8s, (c) YOLOv8m, (d) YOLOv8l, (e) YOLOv8x.

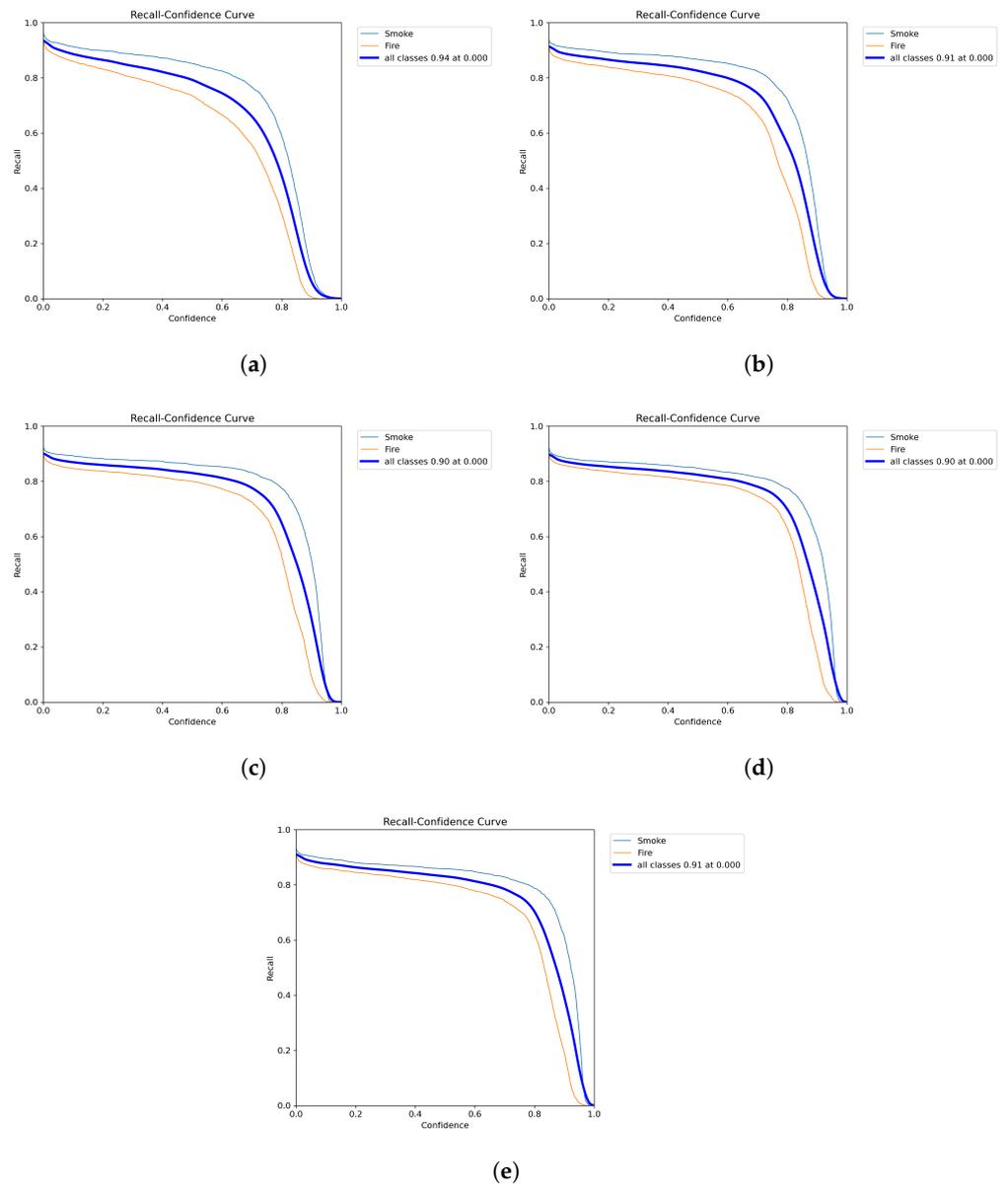


Figure 3. Recall curves of YOLOv8 models: (a) YOLOv8n, (b) YOLO v8s, (c) YOLOv8m, (d) YOLOv8l, (e) YOLOv8x.

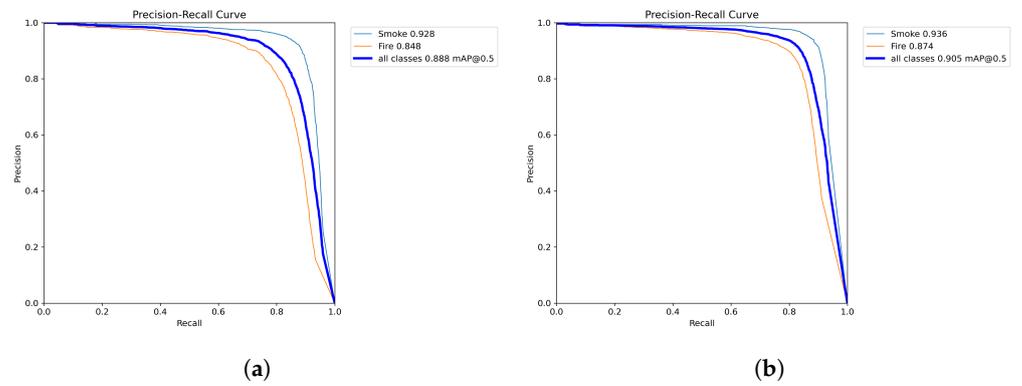


Figure 4. Cont.

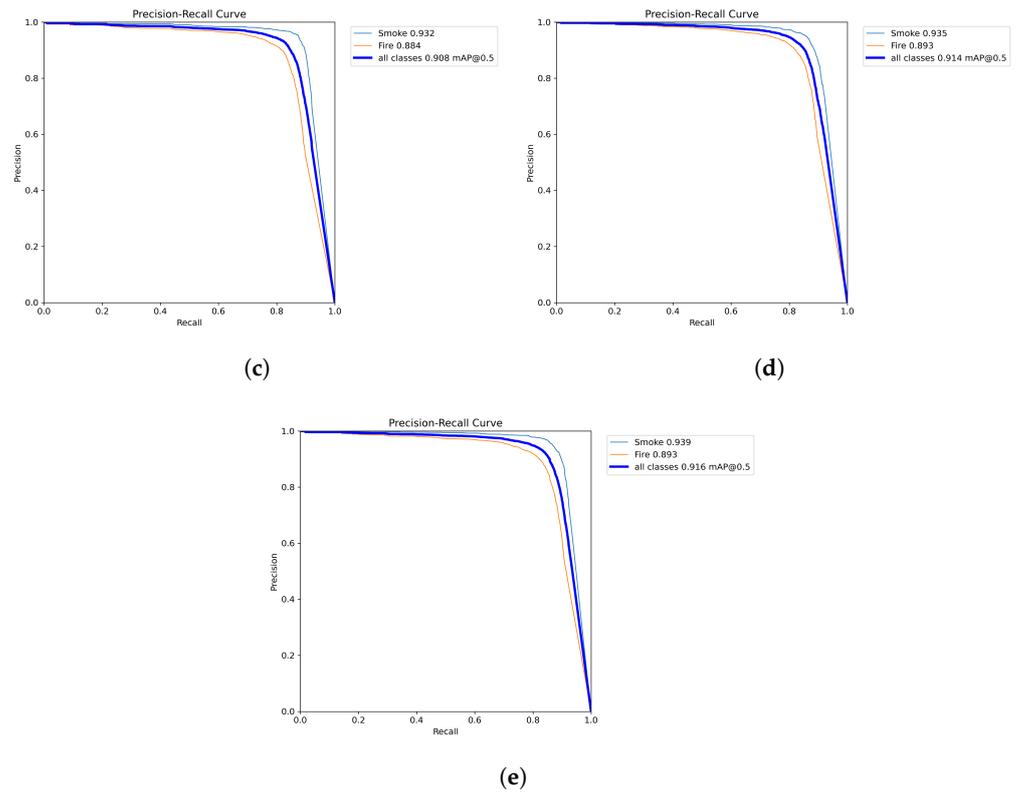


Figure 4. Mean average precision (mAP:50) curves of YOLOv8 models: (a) YOLOv8n, (b) YOLOv8s, (c) YOLOv8m, (d) YOLOv8l, (e) YOLOv8x.

Table 1. Comparative analysis of YOLO models for smoke and fire detection.

Model	Precision	Recall	mAP:50	mAP:50-95
YOLOv8n	0.919	0.793	0.869	0.658
YOLOv8s	0.929	0.828	0.891	0.721
YOLOv8m	0.935	0.831	0.895	0.745
YOLOv8l	0.949	0.837	0.901	0.753
YOLOv8x	0.954	0.848	0.926	0.772
YOLOv7	0.881	0.778	0.854	0.647
YOLOv7-X	0.918	0.817	0.882	0.715
YOLOv7-W6	0.922	0.824	0.887	0.745
YOLOv7-E6	0.937	0.824	0.896	0.748
YOLOv6l	0.582	0.605	0.852	0.496
Faster-RCNN	0.437	0.374	0.471	0.348
DETR	0.443	0.362	0.413	0.291

Figure 5 presents some of the detected fire and smoke using the best model (YOLOv8x) on the test set. As we can see, the model perfectly detects the smoke and fire in different conditions (night, day, forest, street, etc.). For example, in the first images, (a) the model detects the smoke and fire at night time with a high confidence score of 0.83.

For Figure 5d,f,g, the smoke appears with a low contrast and is far; the model provides accurate detection with a confidence score reaching 0.83. In the same way, for Figure 5c,e,h, fire and smoke are perfectly detected in different places (forest and street) with a high confidence score of 0.9.

Table 2 presents some comparative studies with deep learning models for fire and smoke detection. The proposed models outperform the results presented by Yang et al. [39]. The authors used a modified YOLOv5 using the same dataset and same number of images; the proposed models provided improvements of 6%, 3.8%, and 5.8% in terms of precision, recall, and mAP:50, respectively.

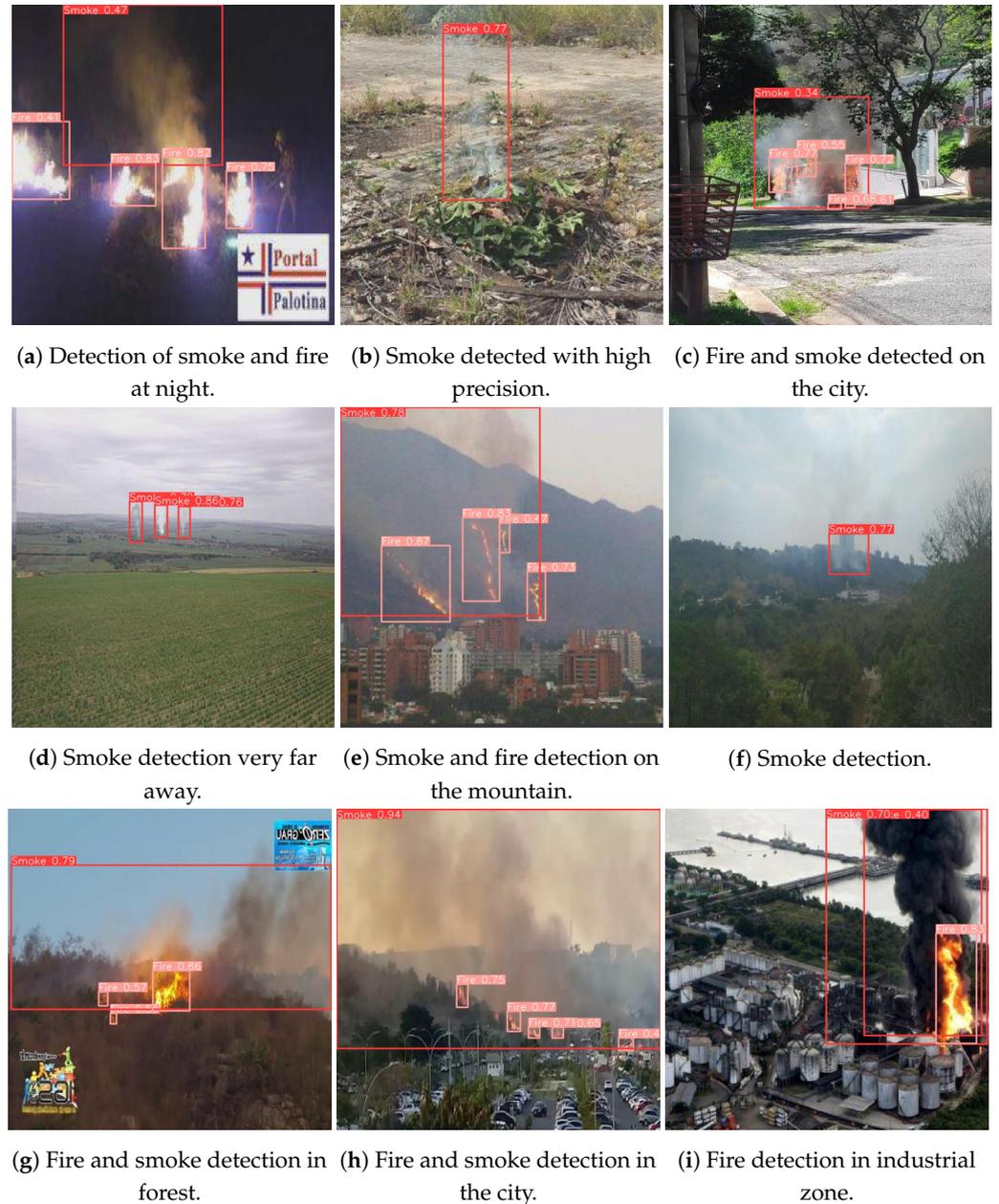


Figure 5. Example of fire and smoke detection by YOLOv8x model.

These qualitative and quantitative results confirm the robustness of the proposed models in the detection of fire and smoke in several locations, distances, and in different periods of the day (day and night).

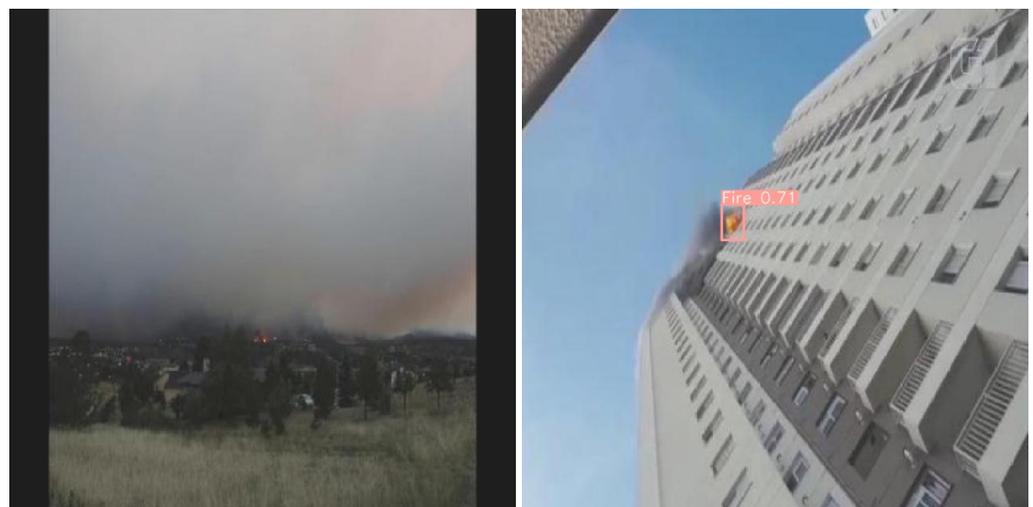
Table 2. Comparison results with related studies.

Study	Model	Precision	Recall	mAP:50	# Images	Detection
Saydirasulovich et al. [41]	YOLOv6	0.934	0.282	-	4000	Fire/Smoke
Talaat et al. [26]	YOLOv8	-	-	0.794	6000	Fire/Smoke
Wei et al. [42]	YOLOv8	-	0.707	0.730	2059	Fire
Xu et al. [43]	YOLOv7	0.861	0.818	0.883	2058	Fire
Yang et al. [39]	YOLOv5	0.892	0.827	0.873	11,667	Fire/Smoke
Proposed model	YOLOv8	0.837	0.952	0.890	11,667	Fire/Smoke

4. Limitation

Detecting fire and smoke poses various challenges. The dynamic and random structure of smoke plumes and the many environmental elements that constitute up the complex forest landscape such as cloud cover and haze complicate the process of detection. It is critical to identify small fires as soon as possible because, if ignored, they may rapidly develop into major catastrophes with fatal consequences. Utilizing computer vision technology to substitute human surveillance provides a highly efficient method of solving these challenges.

Even though the suggested method for detecting smoke from forest fires has proven to be successful, it is crucial to recognize that it has certain limitations. Its notable sensitivity to meteorological factors like haze, clouds, and fog presents an important challenge because these elements can occasionally look similar to smoke (Figure 6a,c). Furthermore, there is another limitation in cases where the shape of smoke is similar to the cloud (see Figure 6b). Our objective is to enhance the model's performance to make accurate predictions by increasing the size of the training dataset and extracting deeper features. An additional solution that could be explored further is the ensemble learning of YOLOv7 and v8 models that can accurately differentiate between the shape and structure of smoke and clouds.



(a) False negative detection of smoke.

(b) False negative detection of smoke

Figure 6. Cont.

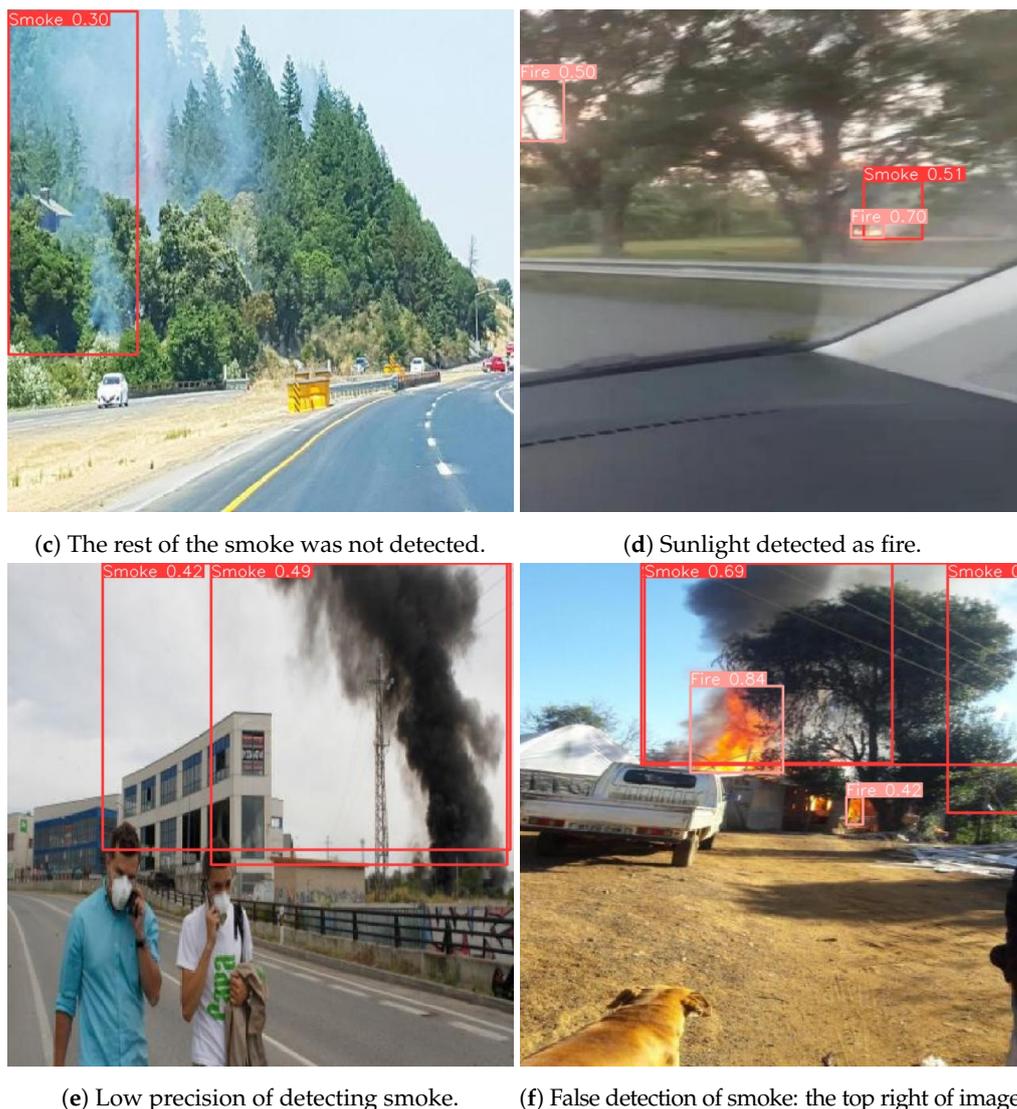


Figure 6. Example of false detection of fire and smoke.

5. Conclusions

In this paper, we presented a fire and smoke detection model based on YOLOv8 on different locations (forest, street, houses, etc.)

YOLOv8 models provided very interesting results compared to YOLOv7 in terms of precision, recall, and mAP:50. All versions of YOLOv8 detected fire and smoke with a high mAP:50 score reaching 0.9 on the day and night time. The scores presented in Table 1 show that YOLOv8 models reached a precision of 0.954, a recall of 0.848, and a mAP:50 of 0.926. We compared our results with the study in [39]. The authors used the same dataset with YOLOv5 with some modification on the network architecture. The proposed models outperformed their obtained scores with improvements of 6%, 3.8%, and 5.8% in terms of precision, recall, and mAP:50, respectively. The developed model can be deployed in several industrial applications such as fire and smoke detection in forests, cities and industrial zones. The proposed model was trained on a large number of fire and smoke images in various locations to enhance the degree of generalization of the model. We aim to integrate the proposed models to be compatible on the edge device for real-time detection, while ensuring optimal performance in demanding environmental conditions.

Author Contributions: Conceptualization, M.C. and M.A.A.; methodology, M.C. and M.A.A.; software, M.C.; validation, M.C. and M.A.A.; formal analysis, M.C. and M.A.A.; investigation, M.A.A. and M.C.; resources, M.C.; data curation, M.C.; writing—original draft preparation, M.C. and M.A.A.; writing—review and editing, M.C. and M.A.A.; supervision, M.A.A.; project administration, M.A.A.; funding acquisition, M.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was enabled in part by support provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2018-06233.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this work come mainly from public datasets [39].

Acknowledgments: Thanks to the support provided by support provided by Calcul Québec (<https://www.calculquebec.ca/>, accessed on 7 April 2024) and the Digital Research Alliance of Canada (alliancecan.ca, accessed on 7 April 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Belavenutti, P.; Chung, W.; Ager, A.A. The economic reality of the forest and fuel management deficit on a fire prone western US national forest. *J. Environ. Manag.* **2021**, *293*, 112825. [CrossRef]
2. Martínez, J.R.M.; Machuca, M.H.; Díaz, R.Z.; y Silva, F.R.; González-Cabán, A. Economic losses to Iberian swine production from forest fires. *For. Policy Econ.* **2011**, *13*, 614–621. [CrossRef]
3. Akter, S.; Grafton, R.Q. Do fires discriminate? Socio-economic disadvantage, wildfire hazard exposure and the Australian 2019–20 ‘Black Summer’ fires. *Clim. Chang.* **2021**, *165*, 53. [CrossRef]
4. Solórzano, A.; Fonollosa, J.; Fernández, L.; Eichmann, J.; Marco, S. Fire detection using a gas sensor array with sensor fusion algorithms. In Proceedings of the 2017 ISOCS/IEEE International Symposium on Olfaction and Electronic Nose (ISOEN), Montreal, QC, Canada, 28–31 May 2017; pp. 1–3.
5. Jang, H.Y.; Hwang, C.H. Methodology for DB construction of input parameters in FDS-based prediction models of smoke detector. *J. Mech. Sci. Technol.* **2020**, *34*, 5327–5337. [CrossRef]
6. Ding, X.; Gao, J. A new intelligent fire color space approach for forest fire detection. *J. Intell. Fuzzy Syst.* **2022**, *42*, 5265–5281. [CrossRef]
7. Emmy Prema, C.; Vinsley, S.; Suresh, S. Efficient flame detection based on static and dynamic texture analysis in forest fire detection. *Fire Technol.* **2018**, *54*, 255–288. [CrossRef]
8. Gao, Y.; Cheng, P. Forest fire smoke detection based on visual smoke root and diffusion model. *Fire Technol.* **2019**, *55*, 1801–1826. [CrossRef]
9. Kim, O.; Kang, D.J. Fire detection system using random forest classification for image sequences of complex background. *Opt. Eng.* **2013**, *52*, 067202–067202. [CrossRef]
10. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
11. Jiang, H.; Learned-Miller, E. Face detection with the faster R-CNN. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 650–657.
12. Kumar, A.; Srivastava, S. Object detection system based on convolution neural networks using single shot multi-box detector. *Procedia Comput. Sci.* **2020**, *171*, 2610–2617. [CrossRef]
13. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
14. Choi, M.; Kim, C.; Oh, H. A video-based SlowFastMTB model for detection of small amounts of smoke from incipient forest fires. *J. Comput. Des. Eng.* **2022**, *9*, 793–804. [CrossRef]
15. Xu, R.; Lin, H.; Lu, K.; Cao, L.; Liu, Y. A forest fire detection system based on ensemble learning. *Forests* **2021**, *12*, 217. [CrossRef]
16. Ghosh, R.; Kumar, A. A hybrid deep learning model by combining convolutional neural network and recurrent neural network to detect forest fire. *Multimed. Tools Appl.* **2022**, *81*, 38643–38660. [CrossRef]
17. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
18. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
19. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.

20. Wang, W.; Xie, E.; Li, X.; Fan, D.P.; Song, K.; Liang, D.; Lu, T.; Luo, P.; Shao, L. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 568–578.
21. Zhan, J.; Hu, Y.; Cai, W.; Zhou, G.; Li, L. PDAM-STPNNet: A small target detection approach for wildland fire smoke through remote sensing images. *Symmetry* **2021**, *13*, 2260. [[CrossRef](#)]
22. Ghali, R.; Akhloufi, M.A.; Jmal, M.; Souidene Mseddi, W.; Attia, R. Wildfire segmentation using deep vision transformers. *Remote Sens.* **2021**, *13*, 3527. [[CrossRef](#)]
23. Barmpoutis, P.; Papaioannou, P.; Dimitropoulos, K.; Grammalidis, N. A review on early forest fire detection systems using optical remote sensing. *Sensors* **2020**, *20*, 6442. [[CrossRef](#)]
24. Saleh, A.; Zulkifley, M.A.; Harun, H.H.; Gaudreault, F.; Davison, I.; Spraggon, M. Forest fire surveillance systems: A review of deep learning methods. *Heliyon* **2024**. [[CrossRef](#)] [[PubMed](#)]
25. Sathishkumar, V.E.; Cho, J.; Subramanian, M.; Naren, O.S. Forest fire and smoke detection using deep learning-based learning without forgetting. *Fire Ecol.* **2023**, *19*, 9. [[CrossRef](#)]
26. Talaat, F.M.; ZainEldin, H. An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Comput. Appl.* **2023**, *35*, 20939–20954. [[CrossRef](#)]
27. Al-Smadi, Y.; Alauthman, M.; Al-Qerem, A.; Aldweesh, A.; Quaddoura, R.; Aburub, F.; Mansour, K.; Alhmiedat, T. Early Wildfire Smoke Detection Using Different YOLO Models. *Machines* **2023**, *11*, 246. [[CrossRef](#)]
28. Wang, Z.; Wu, L.; Li, T.; Shi, P. A smoke detection model based on improved YOLOv5. *Mathematics* **2022**, *10*, 1190. [[CrossRef](#)]
29. Mukhiddinov, M.; Abdusalomov, A.B.; Cho, J. A Wildfire Smoke Detection System Using Unmanned Aerial Vehicle Images Based on the Optimized YOLOv5. *Sensors* **2022**, *22*, 9384. [[CrossRef](#)] [[PubMed](#)]
30. Yazdi, A.; Qin, H.; Jordan, C.B.; Yang, L.; Yan, F. Nemo: An open-source transformer-supercharged benchmark for fine-grained wildfire smoke detection. *Remote Sens.* **2022**, *14*, 3979. [[CrossRef](#)]
31. Abdusalomov, A.; Baratov, N.; Kutlimuratov, A.; Whangbo, T.K. An Improvement of the Fire Detection and Classification Method Using YOLOv3 for Surveillance Systems. *Sensors* **2021**, *21*, 6519. [[CrossRef](#)] [[PubMed](#)]
32. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. *arXiv* **2020**, arXiv:2005.12872.
33. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)] [[PubMed](#)]
35. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
36. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; AAAI: Washington, DC, USA, 2020; Volume 34; pp. 12993–13000.
37. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21002–21012.
38. Feng, C.; Zhong, Y.; Gao, Y.; Scott, M.R.; Huang, W. Tood: Task-aligned one-stage object detection. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 3490–3499.
39. Yang, J.; Zhu, W.; Sun, T.; Ren, X.; Liu, F. Lightweight forest smoke and fire detection algorithm based on improved YOLOv5. *PLoS ONE* **2023**, *18*, e0291359. [[CrossRef](#)]
40. HumanSignal. LabelImg. 2015. Available online: <https://github.com/HumanSignal/labelImg> (accessed on 7 April 2024).
41. Saydirasulovich, S.N.; Mukhiddinov, M.; Djuraev, O.; Abdusalomov, A.; Cho, Y.I. An Improved Wildfire Smoke Detection Based on YOLOv8 and UAV Images. *Sensors* **2023**, *23*, 8374. [[CrossRef](#)] [[PubMed](#)]
42. Wei, Z. Fire detection of YOLOv8 model based on integrated se attention mechanism. *Front. Comput. Intell. Syst.* **2023**, *4*, 28–30. [[CrossRef](#)]
43. Xu, Y.; Li, J.; Zhang, L.; Liu, H.; Zhang, F. CNTCB-YOLOv7: An Effective Forest Fire Detection Model Based on ConvNeXtV2 and CBAM. *Fire* **2024**, *7*, 54. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.