

LEF-YOLO: a lightweight method for intelligent detection of four extreme wildfires based on the YOLO framework

Jianwei Li^{A,*} , Huan Tang^A, Xingdong Li^B, Hongqiang Dou^C and Ru Li^D

For full list of author affiliations and declarations see end of paper

***Correspondence to:**

Jianwei Li
College of Physics and Information
Engineering, Fuzhou University, Fuzhou,
350116, China
Email: lwicq@163.com

Received: 5 April 2023
Accepted: 29 November 2023
Published: 18 December 2023

Cite this:

Li J *et al.* (2024)
International Journal of Wildland Fire **33**,
WF23044.
doi:[10.1071/WF23044](https://doi.org/10.1071/WF23044)

© 2024 The Author(s) (or their employer(s)). Published by CSIRO Publishing on behalf of IAWF. This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License ([CC BY-NC-ND](https://creativecommons.org/licenses/by-nc-nd/4.0/))

OPEN ACCESS

ABSTRACT

Background. Extreme wildfires pose a serious threat to forest vegetation and human life because they spread more rapidly and are more intense than conventional wildfires. Detecting extreme wildfires is challenging due to their visual similarities to traditional fires, and existing models primarily detect the presence or absence of fires without focusing on distinguishing extreme wildfires and providing warnings. **Aims.** To test a system for real time detection of four extreme wildfires. **Methods.** We proposed a novel lightweight model, called LEF-YOLO, based on the YOLOv5 framework. To make the model lightweight, we introduce the bottleneck structure of MobileNetv3 and use depthwise separable convolution instead of conventional convolution. To improve the model's detection accuracy, we apply a multiscale feature fusion strategy and use a Coordinate Attention and Spatial Pyramid Pooling-Fast block to enhance feature extraction. **Key results.** The LEF-YOLO model outperformed the comparison model on the extreme wildfire dataset we constructed, with our model having excellent performance of 2.7 GFLOPs, 61 FPS and 87.9% mAP. **Conclusions.** The detection speed and accuracy of LEF-YOLO can be utilised for the real-time detection of four extreme wildfires in forest fire scenes. **Implications.** The system can facilitate fire control decision-making and foster the intersection between fire science and computer science.

Keywords: convolutional neural networks, deep learning, extreme wildfire, fire safety, lightweight, multiscale feature fusion, object detection, YOLO (LEF-YOLO).

Introduction

Extreme wildfires are large-scale, fast spreading, high-intensity wildfires that often cause great damage (Castro *et al.* 2021). When extreme wildfires occur, they often manifest as different types such as firelines merging, spot fire, crown fire, eruptive fire, fire whirl, conflagration, jump fire, firestorm, and so on (Liu *et al.* 2021). Different types of extreme wildfires may transform into each other under the influence of factors such as wind, topography, and the presence of combustibles (Li *et al.* 2018). Extreme wildfires often produce abrupt changes in fire behaviour, including unpredictable changes in fire intensity, erratic rates and directions of spread, spotting, and winds caused by the occurrence of fire. These changes can pose a significant threat to firefighters and can undo efforts to extinguish the fire (Tedim *et al.* 2018). Moreover, the characteristics of different types of extreme wildfires vary. Hence, to maximise escape time and make better decisions, firefighters will have to distinguish the type of extreme wildfires and locate the location of the extreme wildfires in a timely manner (Viegas and Simeoni 2011; Colston and Flik 2012). The NWCG (National Wildfire Coordinating Group 2020) notes that high diffusion rates, numerous occurrences of crowning, spotting, fire rotation, and strong columns of convection are often present in extreme wildfires, further complicating predictability. While extreme wildfires occur less frequently, they pose a greater hazard, with only 3% of uncontrolled fires causing 95% of the damage (Liu *et al.* 2021). Unfortunately, tragedies resulting from extreme wildfires are not uncommon. For example, the Yarnell Hill Fire in central Arizona, USA in June 2013 claimed the lives of 19

trained firefighters who had deployed to a fire shelter (Department of Forestry and Fire Management 2013). In March 2020, 19 people lost their lives, and three others were injured while fighting a forest fire on the Pijiashan ridge in Liangshan Prefecture, China due to sudden changes in wind direction, flying fire breaks, and self-rescue failure. The fire resulted in over 3000 ha of total land overfire area, with nearly 791 ha of forest affected and economic losses of approximately CNY100 million (The People's Government of Sichuan Province 2020). Extreme wildfires represent dangerous natural disasters that cause significant damage to both human life and property. If they can be detected automatically and accurately in real time, it will give firefighters more time to make decisions.

In this context, computer vision-based target detection technology is of great importance (Li *et al.* 2018; Lei *et al.* 2023). This can help firefighters determine wildfire status and trends faster and more accurately thus providing real time early warning information (Wu *et al.* 2023). Target detection technology can use high-definition cameras to capture images and videos of fires and use advanced algorithms to automatically identify and analyse fires. In addition, it can be applied to fireground mobility equipment and fire warning robots to improve firefighter efficiency and safety. However, it is not the case that a larger and more complex target detection algorithm model will be more advantageous in extreme wildfire detection tasks (Zaidi *et al.* 2022). First, extreme wildfire scenes are usually resource-constrained and hence fire detection on embedded or mobile devices requires consideration of the device's limited computing power and storage resources. Second, lightweight models usually have faster inference speed, which enables fast detection and response in near real time scenarios and improves fire processing efficiency. In addition, lightweight models are more adaptable to mobile deployment needs and have higher flexibility and mobility. Finally, detection results sometimes need to be transmitted over the network to other mobile devices or servers for subsequent processing or raising alarms. Lightweight models can reduce the size of the model and reduce the latency and bandwidth requirements for network transmission. Therefore, a lightweight target detection algorithm model with excellent performance has important advantages in extreme wildfire detection tasks. Such an algorithm can improve the efficiency and safety of extreme wildfire processing and provide strong support for firefighters' work.

Related work

The study of extreme wildfires has received increasing attention in recent years. Tedim *et al.* (2018) proposed a classification of wildfires based on measurable data and behavioural parameters of fire spread and gave rules for classifying extreme fire events. Some scholars investigated

extreme fire generation mechanisms using physics-related knowledge and mathematical models. Tohidi *et al.* (2018) used knowledge of fluid dynamics to review the conditions and structure of fire cyclone formation in extreme fire types. Gómez-Vázquez *et al.* (2014) developed density management diagrams in an attempt to analyse canopy fire potential using mathematical models. Liu *et al.* (2021) used knowledge of combustion dynamics to explore the causes and patterns of small-scale flame developing into extreme fires and gave an overview of the interconversion relationships between the various types of combustion dynamics to investigate the causes and patterns of small-scale fires expanding into extreme fires. They also gave the interconversion relationships between each type of extreme fire. The construction of mathematical models is indeed useful for analysing the conditions of extreme fire formation, but this approach is difficult to apply to realistic and complex extreme fire scenarios. It also does not give rapid real-time fire scene information.

With the development of computational vision technology and machine learning (Chen 2022; Sharma *et al.* 2022), researchers have aimed to perform flame detection using image processing-related techniques. Pritam and Dewan (2017) focused on flame colour features and performed flame detection by combining LUV colour space with a hybrid transform. Wang *et al.* (2017) proposed a flame detection model based on flame kinematic features. Chen *et al.* (2022) used a multimodal UAV to collect RGB and IR fire images to detect flames through a combination of multimodal approaches. Traditional flame detection methods focus on the colour, shape and motion features specific to flames. Hence image processing using these features requires high demands on the quality of the images.

The rapid development of deep learning has enabled researchers to detect wildfires in more complex scenarios (Qiang *et al.* 2021; Azim *et al.* 2022). YOLOv5 and EfficientDet were used to detect wildfires in different scenes (Muhammad *et al.* 2019), and Sudhakar *et al.* (2020) introduced an FFD-compatible multidrone wildfire detection system. Barmpoutis *et al.* (2019) used R-CNN to perform multidimensional texture analysis by combining deep learning with multidimensional texture analysis and using linear dynamical systems and VLAD coding for texture analysis techniques to achieve flame detection tasks. Substantial progress has been made via these methods for the identification of conventional wildfires, but the identification of extreme wildfires requires a more fine-grained analysis and methods that only identify the presence or absence of wildfires cannot be applied to identify extreme wildfires.

Deep learning models typically produce complex and large models. To facilitate deployment on mobile devices, some researchers have already conducted lightweight processing on deep neural networks in many fields (Bao *et al.* 2021; Gonzalez-Huitron *et al.* 2021; Diao *et al.* 2022; Cui *et al.* 2023). Similarly, a lightweight model is also required

for wildfire scenes. Almeida *et al.* (2022) proposed an edge smoke flame detection model that combines edge devices with convolutional networks by removing redundant convolutional layers to enable real time fire detection. Shees *et al.* (2023) proposed FireNetv2, which is much less computationally intensive and trainable with only 318 460 parameters and can be used to piggyback on embedded devices for fire detection.

In this paper, we propose an extreme wildfire detection model based on lightweight deep learning. Instead of relying on the physicochemical properties of extreme wildfires, we analyse and process video images of extreme wildfires using a combination of deep learning and computer vision techniques to detect them. Existing wildfire detection methods can only distinguish between the presence or absence of wildfires and cannot be applied to classify extreme wildfires, which may have different shapes, behaviour patterns, and training scenarios than conventional wildfires. As shown in Fig. 1, our proposed method can detect different types of common extreme wildfires, including firelines merging, fire whirl, crown fire, and spot fire. It should be noted that for crown fires and fire whirls, our research work detects the entire fire spread process, while in the detection of firelines merging and spot fires, detection is performed for some stages and not the entire process. This depends on the specificity of the spread process of different categories of extreme wildfires. For firelines merging, the process of firelines merging is the process of two or more fire lines merging from far to near, from completely non-intersecting to intersecting and finally to

completely merging together. We detect the stage of intersection or convergence in firelines merging. Spot fires follow three sequences: (1) generation of firebrands; (2) loading, transport and settling of firebrands; and (3) ignition of receptor fuels by firebrands (spot fire ignition). Spot fires develop in more distinct phases. This is because the first two phases of spot fires produce a large number of rising or floating firebrands, which show significant scattered features on the captured wildfire images. Computer vision can combine these features with background information for spot fires detection. The tendency of firelines merging and spot fires to form siege more easily creates great difficulties for firefighters to escape and endangers the lives of firefighters. The significance of our research work is that it detects extreme wildfire hazards that may appear early so that firefighters can have more time to make the right decisions and protect their lives.

Materials and methods

Extreme wildfire dataset

Extreme wildfires are diverse in category, and we focus on four common types: crown fires, spot fires, firelines merging, and fire whirls. They exhibit distinct characteristics, formation factors, and potential transformation sources. Table 1 summarises and compares these four types of extreme wildfires (Werth *et al.* 2011, 2016; Liu *et al.* 2021; Hantson *et al.* 2022). Fig. 2 shows the potential

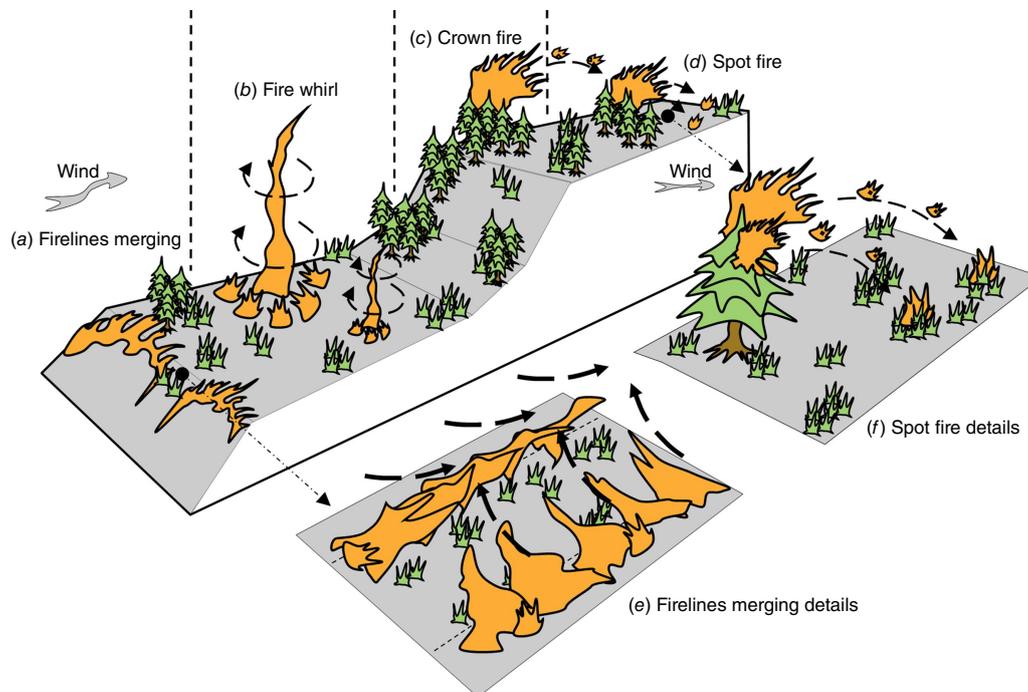


Fig. 1. Visual schematic of four types of extreme wildfires. (a) Firelines merging, (b) fire whirl, (c) crown fire, (d) spot fire, (e) firelines merging details, and (f) spot fire details.

Table 1. Description, characteristics, formation factors and potential sources of transformation of four types of extreme wildfires.

		Category of extreme wildfires			
		Crown fire	Firelines merging	Spot fire	Fire whirl
Description		Fires that burn into the vegetation crown and spread rapidly along the crown.	A fire in which the fire lines of two or more fires merge to form a larger fire.	Fires caused by sparks, sparks or burning material flying around the fire source into the surrounding area.	Fires formed by local rotation of the atmosphere, where flames and smoke form a vortex in the rotating air stream.
Characteristic	Flame shape	Conical, crown-shaped	Irregular, mostly serrated	Scattered	Conical, striped
	Fire intensity	High	High	Low	High
	Spread rate	Fast	Relatively fast	Relatively slow	Fast
	Temperature	High	High	Moderate	High
	Flame height	Large	Moderate	Low	Large
	Range	Wide	Relatively Large	Small	Relatively Small
	Continuity	Continuous	Continuous	Discontinuous	Continuous
Formation factors	Terrain	Steep	–	Moderately steep	–
	Wind	High	Low	High	High
	Humidity	Dry	Moderately dry	Very dry	Moderately dry
	Fuel	Dense trees	Low vegetation, dry grass	Light-weight tree trunks, branches, leaves	Fallen leaves, debris, branches, dry grass
Potential sources of transformation	Firelines merging	Spot fire	Fire whirl, Crown fire	Firelines merging	

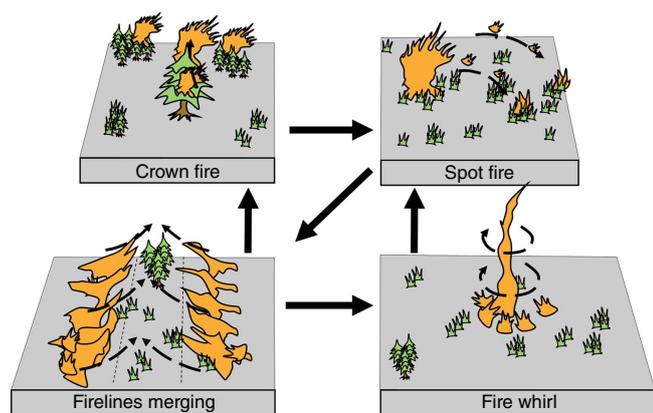


Fig. 2. Transition between four types of extreme wildfires.

transformation relationships among them. Models should detect extreme wildfires in transformation along with extreme fire categories before and after transformation. Our study employs a combination of computer vision and deep learning techniques to recognise these four types of extreme wildfires. After collecting, annotating, and dividing the extreme wildfire image dataset, we train an extreme wildfire detection model.

In this study, a specialised dataset was constructed for the extreme wildfire fire detection task. It contains a large amount of images and annotation information related to extreme wildfires. There are 2330 crown fire images, 1422 firelines

merging images, 815 spot fire images and 1247 fire whirl images in the dataset, totalling 5814 images. The dataset was divided into a training set of 4789 images, a validation set of 516 images and a test set of 509 images. Examples of the dataset samples are in Fig. 3. The process of dataset construction involved three steps: (1) data acquisition; (2) annotation information; (3) and data augmentation. For data acquisition, image data related to extreme wildfires were collected through various channels and sources, including fire experiments, surveillance videos, image search engines, etc. The data collection process covered multiple scenarios, different times of day, and various fire situations. For labelling information of each extreme wildfire image, we used the Labelling tool to label detailed information, including the bounding box coordinates and category labels of the fire area. During the annotation process, professionals were used to annotate the images, and quality control and validation were performed to ensure the accuracy and consistency of the annotation results. For data augmentation, appropriate data augmentation techniques can increase the number of data samples, reduce overfitting of the network model, and improve the generalisability of the model application. The data augmentation methods used in this study included horizontal and vertical flips, Gaussian blur, sharpening, affine transformation, cropping and scaling, random deformation using control points, contrast augmentation (with random values between 0.75 and 1.5 applied to each channel), introduction of Gaussian noise,



Fig. 3. Selected images of the extreme wildfire dataset. (a) Crown fire, (b) firelines merging, (c) spot fire, and (d) fire whirl.

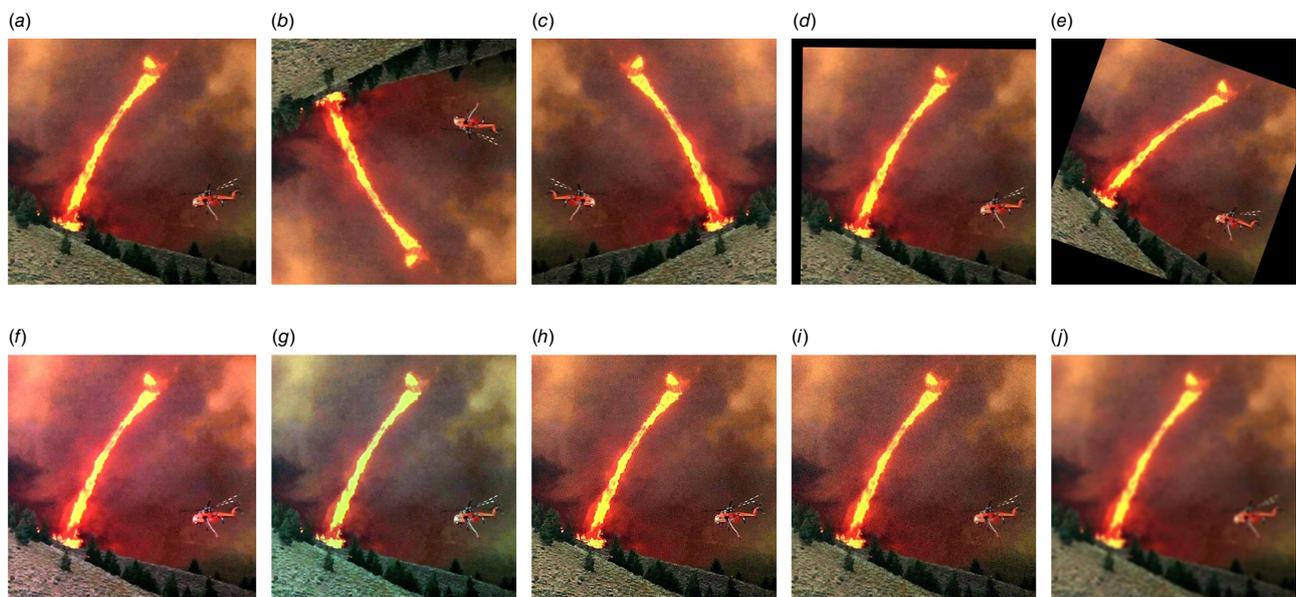


Fig. 4. Data augmentation methods. (a) Original image, (b) vertical flip, (c) horizontal flip, (d) affine transformation, (e) random rotation, (f) contrast augmentation, (g) colour channel change, (h) sharpening, (i) adding noise, and (j) blurring.

and changes in brightness or colour. These methods are in Fig. 4. To aid the research and developer community, we have indicated in Data availability that this dataset will be shared upon reasonable requests.

Model improvements

YOLOv5

YOLO (You Only Look Once) is a one-stage target detection algorithm that identifies objects and locations in images with only one view (Redmon *et al.* 2016). It has been widely used in agriculture (Dang *et al.* 2023), security (Qin *et al.* 2022), and medicine (Wu *et al.* 2021). Many researchers have improved YOLO to accommodate more specific needs (Nguyen *et al.* 2019; Sadykova *et al.* 2020; Hsu and Lin 2021; Dai *et al.* 2022; Qin *et al.* 2022). YOLOv5 (Jocher

2020), a member of the YOLO family, has a faster detection speed and higher detection accuracy than YOLOv3 and YOLOv4 and a more flexible architecture that can adapt to different tasks. YOLOv5 uses CSP (cross-stage partial) (Wang *et al.* 2020) as its backbone part. The spatial pyramid pooling (SPP) block structure is used before the neck structure to increase the perceptual field. It uses a PAN (path aggregation network) structure based on the FPN (feature pyramid network) structure (Lin *et al.* 2017) as its neck part. The head part of YOLOv5 is the same as that YOLOv3 and YOLOv4, which ultimately generate three scales of feature maps, thus enabling multiscale detection. YOLOv5 has five versions (YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x) according to different network depths and widths. YOLOv5 is the classic version of the YOLO series, a model that has been widely used and validated. It has

achieved good results in many applications. For this study, YOLOv5s can meet the requirements of detecting extreme wildfires and can also make many lightweight improvements. Hence we chose YOLOv5s as the baseline model for constructing our improved model.

Lightweight network model Mobilenetv3

Mobilenetv3 is a lightweight neural network model that achieves high accuracy by combining hardware-aware NAS (Network Architecture Search) with the Net-Adapt algorithm (Howard et al. 2019). The network architecture of Mobilenetv3 is in Fig. 5. The model's lightness and high accuracy are achieved by using depthwise separable convolutions combined with the inverted residual with a linear bottleneck approach. The core of Mobilenetv3 lies in its use of depthwise separable convolutions, which are composed of DW (depthwise) and PW (pointwise) convolutions. Compared to conventional convolutional operations, depthwise separable convolutions significantly reduce the number of parameters and computational cost.

MobileNetv3 utilizes the H-sigmoid activation function in its attention mechanism, which offers faster computation than the sigmoid function and resolves the issue of gradient vanishing (Eqn 1). Within the bottleneck, the H-swish activation function is used to replace the previous ReLU6 activation function. The H-swish expression (Eqn 2) offers faster and smoother computation on CPUs, ensuring neural network stability and achieving higher accuracy in deep learning image classification tasks.

$$H_sigmoid(x) = \max\left(0, \min\left(1, \frac{x + 1}{2}\right)\right) \quad (1)$$

$$H_swish(x) = x \frac{ReLU6(x + 3)}{6} \quad (2)$$

In our study, we applied the relevant structure of Mobilenetv3 to our model while considering both the model size and detection accuracy.

Coordinate attention module

CA (coordinate attention) is a new and efficient lightweight attention mechanism module for neural networks, as

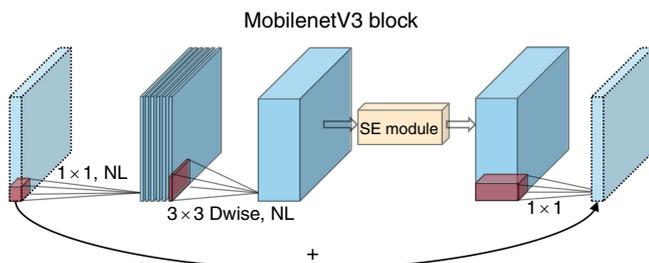


Fig. 5. Network structure of Mobilenetv3.

proposed by Hou et al. (2021). Compared with existing attention mechanisms such as squeeze-and-excitation (SE) and convolutional block attention module (CBAM), the CA module considers both interchannel information and location information. The CA structure is shown in Fig. 6. The CA module splits channel attention into two parallel 1D feature encodings to avoid the loss of positional information resulting from direct 2D global pooling. This results in attention maps with spatial coordinate information. Specifically, averaging the inputs along the horizontal and vertical directions yields two separate direction-aware feature maps, namely, $(C \times H \times 1)$ and $(C \times 1 \times W)$. These two feature maps are then embedded with vertical direction information and horizontal direction information. These are then encoded into two attention maps, and the position information with long-range dependence can thus be stored in the generated attention maps. As a result, CA with channel information and direction-aware and position-sensitive information can locate and identify the target area more accurately.

The proposed LEF-YOLO architecture for extreme wildfire detection

We now present the proposed LEF-YOLO for the detection of four extreme wildfires. Our model should be deployable on

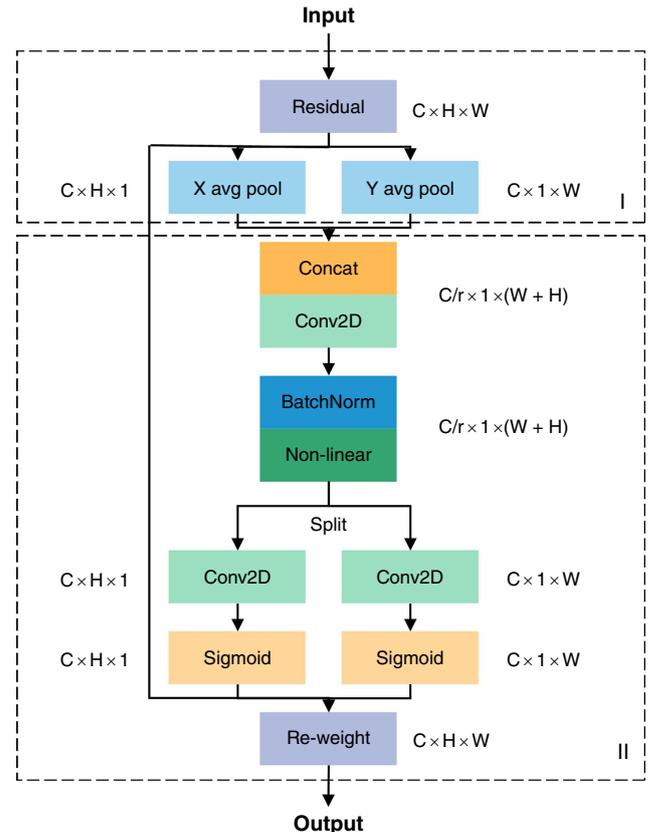


Fig. 6. Network structure of Coordinate Attention (CA).

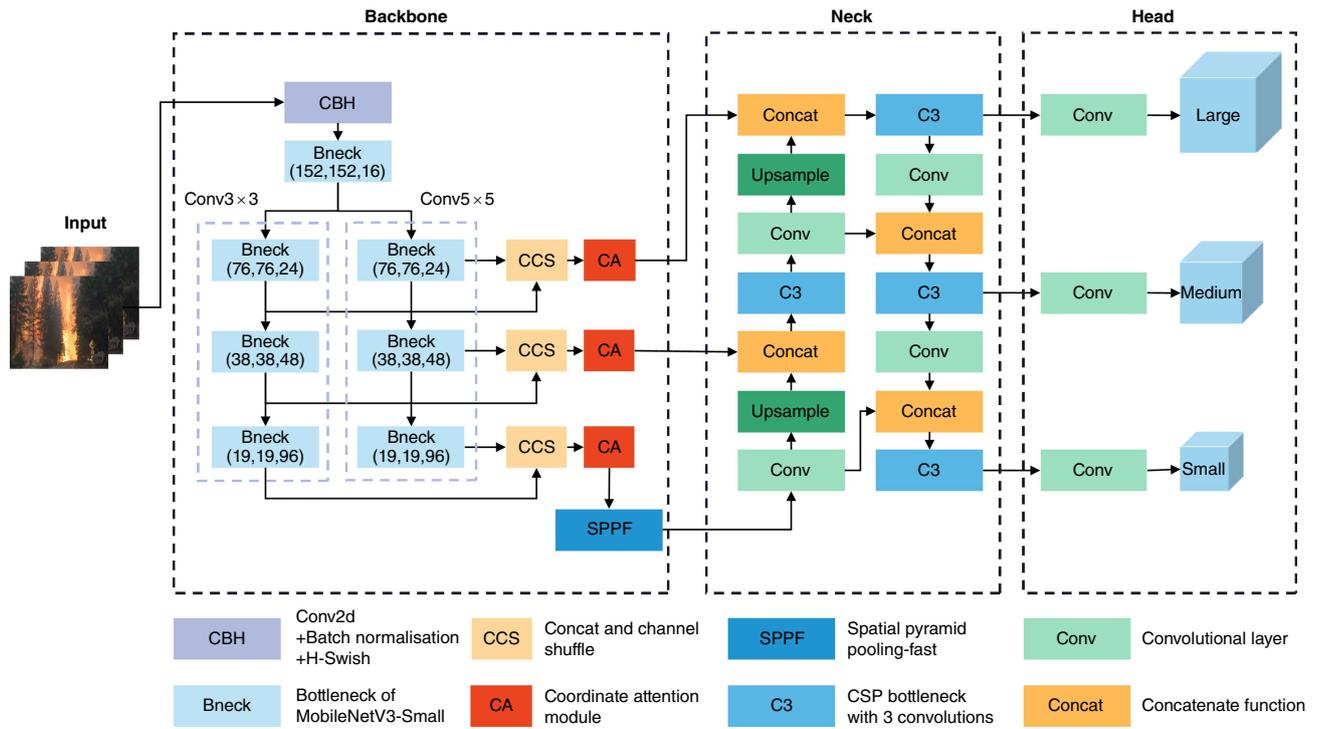


Fig. 7. LEF-YOLO network structure.

embedded or mobile devices. This requires lightweight models that can accommodate resource constraints, improve processing efficiency and reduce latency and bandwidth requirements for network transmission. Fig. 7 illustrates the structure of our proposed LEF-YOLO framework, which consists of three main parts: the backbone, neck, and head. Further details of LEF-YOLO are described below.

To improve the model’s ability to extract extreme wildfire features while reducing the complexity of the model, we propose a new backbone to replace the original CSP of YOLOv5. The new backbone uses a two-branch, three-level structure. The two branches correspond to the use of two depthwise convolutional kernels of different sizes. The 3×3 convolutional kernel is used to capture the local details and edge features of the extreme wildfire image, while the 5×5 convolutional kernel has a larger perceptual field and better captures the overall shape and structure of the flame. The combination of the two results in a more comprehensive and feature-rich representation capability. The combination of the two will be beneficial for the model to extract richer flame features such as conical, striped features for fire whirls, scattered features for spot fires, and so on. Our model downsamples the feature image multiple times by the bottleneck structure in Mobilenetv3-small, which actually contains multiple depthwise separable convolutions in one bottleneck structure. Each downsampling increases the perceptual field, and by doing so, more levels of abstract feature representations at different scales can be obtained, which facilitates effective detection of

extreme wildfire targets at different scales. Earlier downsampling layers can capture low-level image features, such as edges and textures, while later downsampling layers can learn higher-level semantic features, such as the shape and structure of the flame. The separable convolution with sufficient depth allows the model to capture rich extreme wildfire features and contextual information, including the environment, such as the sky and the terrain around the flames, and the smoke and burning vegetation associated with extreme wildfire intensity.

Specifically, the input 608×608 pixels image is first transformed into a 304×304 feature map using CBH (Conv2d, Batch Normalisation, and H-swish) and then into a (152, 152, 16) feature map using Bottleneck. This feature map is then fed into the depthwise separable convolution with 3×3 and 5×5 kernels (Fig. 8). After the depthwise separable convolution of different sizes, we obtain feature maps of three different sizes: (76, 76, 24); (38, 38, 48); and (19, 19, 96). These feature maps are then fused with the same channel in two according to their corresponding sizes and fed into the CA module. Feature fusion uses concat and channel shuffle structures that reduce computational complexity (Ma et al. 2018). The CA module obtains interchannel information and direction-related position information thereby helping the model to accurately locate extreme wildfire feature targets.

The neck further processes the 76×76 and 38×38 feature maps, while the 19×19 feature map undergoes spatial pyramid pooling-fast (SPPF) before being input to the FPN + PAN structure. The FPN upsamples the feature

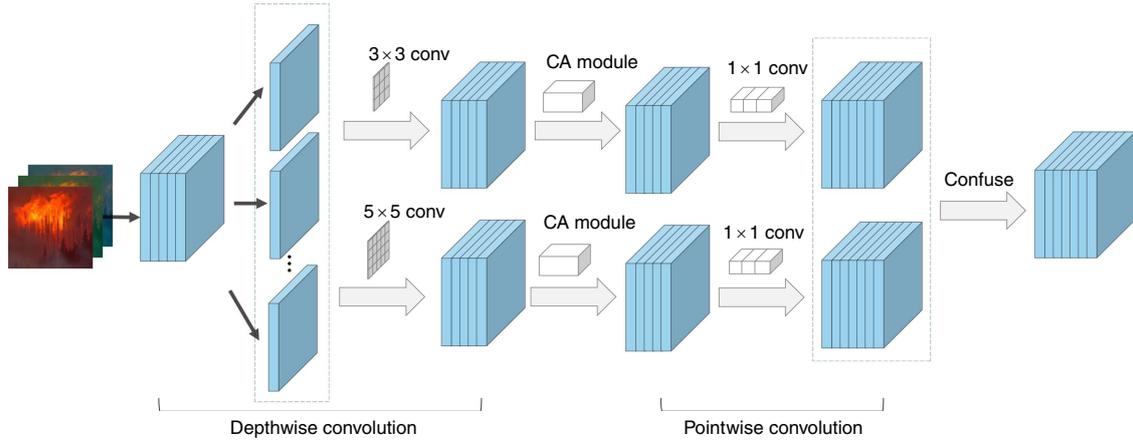


Fig. 8. Feature fusion after multi-size depthwise separable convolution.

maps so that the 76×76 and 38×38 feature maps contain stronger semantic information. PAN downsamples the feature maps so that the 19×19 and 38×38 feature maps contain stronger location information related to extreme wildfires. The two features are fused so that all three sizes of feature maps contain stronger semantic and feature information thus ensuring the model's accurate prediction.

The proposed LEF-YOLO model predicts three different scales of bounding boxes: (1), 76×76 ; (2) 38×38 ; and (3) 19×19 , and classifies the target classes to detect extreme wildfires.

The loss function of LEF-YOLO is composed of regression loss, confidence loss, and classification loss, which is identical to YOLOv5. BCEWithLogitsLoss is still selected as the confidence and classification loss functions in LEF-YOLO. Moreover, based on experimental results, CIoU (complete intersection over union), which performs better than GIoU (generalised intersection over union), is adopted as the loss function for bounding box regression. CIoU takes into account the aspect ratio of the bounding box, and the expression of the CIoU loss function is given in Eqns 3–6.

$$\text{CIoU} = \text{IoU} - \left(\frac{d^2}{c^2} + \alpha v \right) \quad (3)$$

$$v = \frac{4}{\pi^2} (\arctan R^{\text{gt}} - \arctan R)^2 \quad (4)$$

$$\alpha = \frac{v}{(1 - \text{IoU}) + v} \quad (5)$$

$$L_{\text{CIoU}} = 1 - \text{CIoU} \quad (6)$$

The intersection over union (IoU) represents the overlapping area between the predicted box and the target box; d represents the distance between the centres of the predicted and target boxes; while c represents the distance between the diagonal points of their minimum bounding rectangles; and R and R^{gt} represent the aspect ratios of the predicted and target boxes, respectively.

Evaluation metrics

Our study used precision, recall, $F1$ -score and mAP as evaluation metrics to validate the model. Precision represents the percentage of detection results that are true fires. Recall represents the proportion of all real fires that are successfully detected. The $F1$ -score balances accuracy and comprehensiveness by considering a combination of precision and recall evaluation metrics. The mAP (mean average precision) represents the average performance of fire detection models for different classes of fires under their respective precision-recall curves. The reliability of these metrics may be limited in the absence of data, but they can still provide some useful information about model performance and overall performance. The definitions of precision, recall, $F1$ -score and mAP are in Eqns 7–10. The precision-recall curve (P-R curve) combines the precision and recall of the model. Usually, P-R curves are used to address unbalanced category distributions or lack of positive samples in the dataset and can help us observe the trend between the precision and recall of the model under different thresholds.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

$$F1_{\text{score}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

$$\text{mAp} = \frac{\sum_{i=1}^n \text{AP}(i)}{n} \quad (10)$$

where TP is the number of true positives; TN is the number of true negatives; FN is the number of false-negatives; FP is the number of false-positives; AP is the average precision of the current species, and n is the number of species. AP is the average precision of the current species, and n is the number of species.

Results and discussion

To verify the effectiveness of our proposed model, we conducted a number of experiments. Fig. 9 shows the variation in the loss value with the number of training iterations in

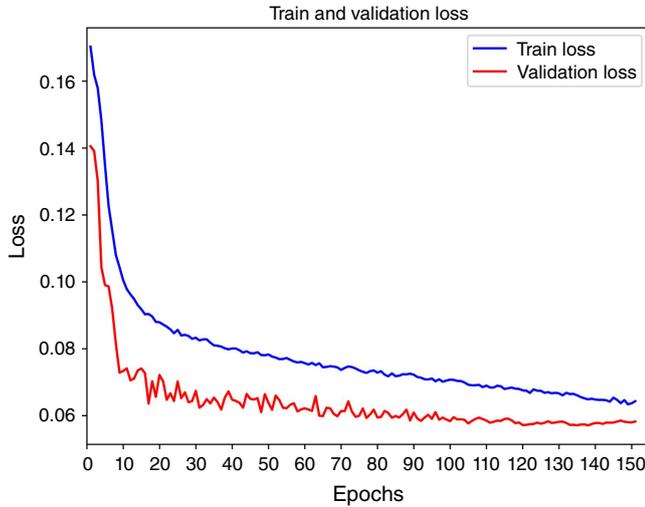


Fig. 9. Training set loss curves and validation loss curves of the proposed model on the extreme wildfire dataset.

the training and validation phases. The loss value gradually decreases as the training proceeds. This indicates that the model is gradually optimised through iterations to learn better feature representation and pattern detection capabilities. In the training set, we applied data augmentation techniques to introduce noise and ambiguity. This making the model more robust to different types of inputs. Since the validation set consisted of clean data, the validation set loss was slightly lower than that of the training set loss. However, the validation loss is not the only metric to evaluate the model and instead, the following is a more comprehensive evaluation of the model. Fig. 10 shows partial results of LEF-YOLO for the detection of four extreme wildfires in real time. We conducted several sets of experiments on a homemade extreme wildfire dataset to compare and analyse the performance of our proposed algorithm with various state-of-the-art algorithms.

Performance comparison between our proposed algorithm and state-of-the-art algorithms

LEF-YOLO was compared with five other state-of-the-art algorithms, including SSD-300 (Liu *et al.* 2016), YOLOv3, YOLOv4s-mish, and YOLOv5s. We discuss the model

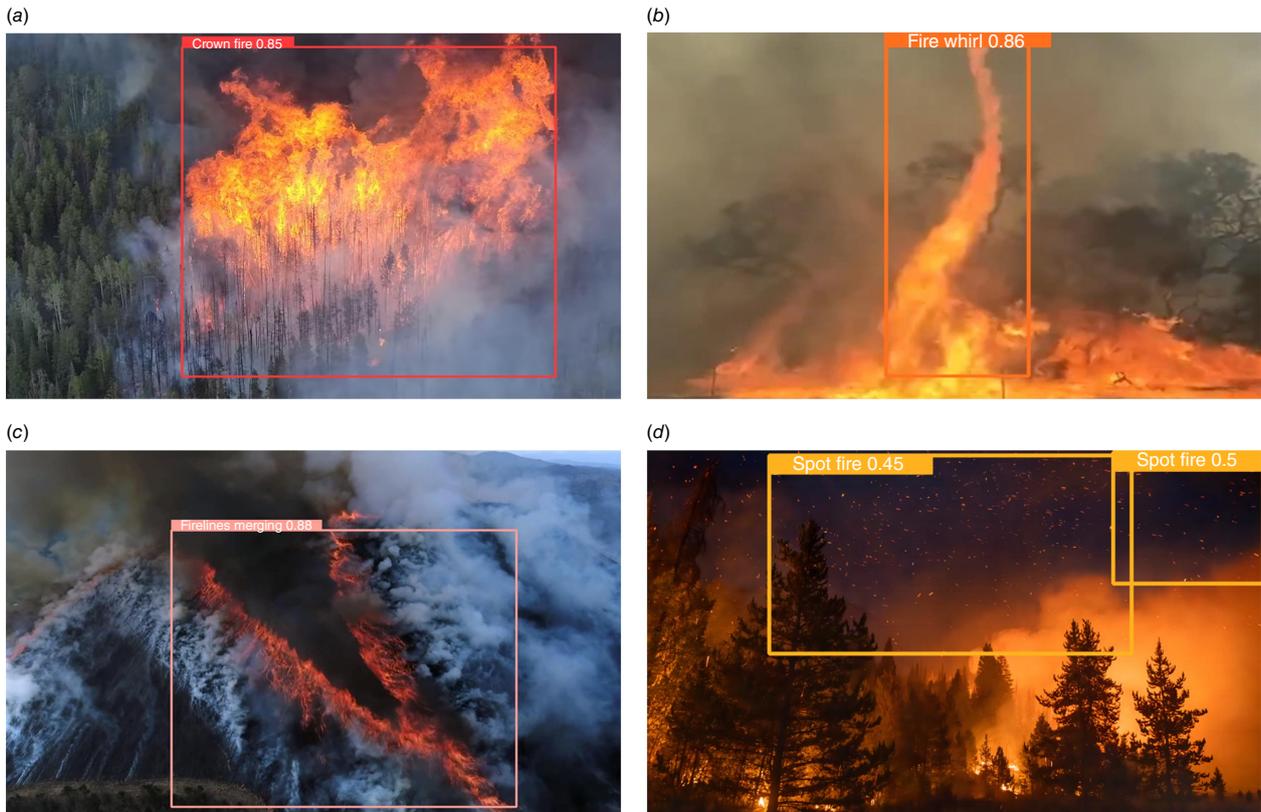


Fig. 10. Partial detection results of four extreme wildfires by LEF-YOLO. (a) Crown fire, (b) fire whirl, (c) firelines merging, and (d) spot fire.

performance for both model detection accuracy and model detection speed.

Comparison of the detection accuracy between our proposed algorithm and the state-of-the-art algorithm

In terms of model detection accuracy, the experimental results (Fig. 11, Table 2) divide extreme wildfires into four classes with different average % accuracies: (1) crown fires (76.7%); (2) firelines merging (95.4%); (3) fire whirls (93.2%); and (4) spot fires (86.0%). The results show that our model can achieve high detection accuracy for all four types of extreme wildfires. The average accuracy of all

other extreme wildfire classes is over 80% except for crown fire, which shows lower accuracy. LEF-YOLO outperforms the one-stage identification SSD-300, with crown fire, firelines merging, and fire whirl showing increased average precision (AP) of 11.7, 20.7, and 33.4%, respectively. Additionally, LEF-YOLO shows higher mAP and F1-score by 24.8 and 26%, respectively. The YOLOv3, YOLOv4s-mish, and LEF-YOLO models also perform well against each other, with a 19.3 and 7.9% increase in mAP and a 15 and 9% increase in F1-score, respectively. Compared to the basic YOLOv5s model, the F1-score is 82%, while the mAP is slightly higher by 4.6%. Finally, LEF-YOLO has the highest mAP and F1-score while maintaining a high AP for all types of extreme wildfires. It has the smallest computational effort and model size while

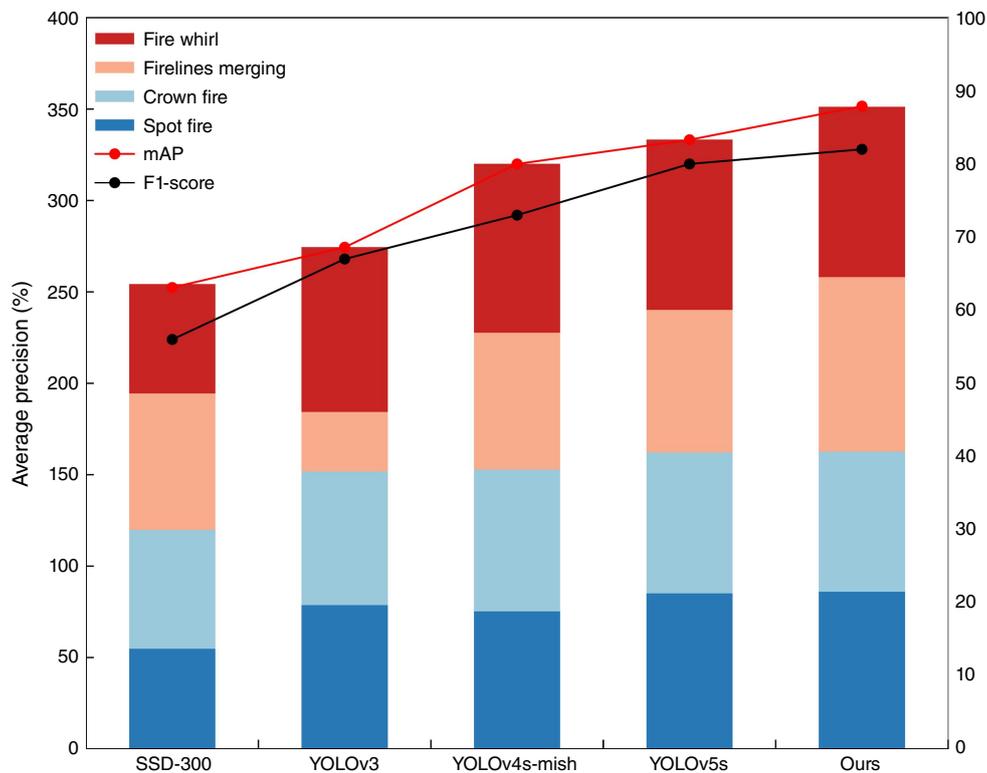


Fig. 11. Comparison of Average Precision between different algorithms for various types of extreme wildfires; the line shows the comparison of mAP and F1-score between different algorithms.

Table 2. Comparison between different algorithms for each type of extreme wildfire.

Algorithms	AP (%)				mAP (%)	F1 score (%)
	Crown fire	Firelines merging	Fire whirl	Spot fire		
SSD-300	65.0	74.7	59.8	54.8	63.1	56
YOLOv3	73.1	32.5	90.2	78.7	68.6	67
YOLOv4s-mish	77.5	74.9	92.3	75.3	80.0	73
YOLOv5s	77.0	78.0	93.2	85.1	83.3	80
Ours	76.7	95.4	93.2	86.0	87.9	82

maintaining the second highest precision and the highest recall. The model has a 61 frames per second image detection speed in the GPU and the shortest inference time in the CPU, which is more advantageous when the model is deployed in resource-limited mobile or embedded devices. The above comparison experiments show that our model has better performance in aggregate. As shown in the line graph (Fig. 11), LEF-YOLO outperforms other algorithmic models with higher evaluation metrics of mAP and F1-score.

In Fig. 12a, l demonstrates that in the extreme wildfire identification process, SSD-300 and YOLOv5s miss detection, while LEF-YOLO accurately marks the extreme wildfire target. In contrast, LEF-YOLO has a high confidence level in identifying the fire whirl target in the second column of Fig. 12. The third column of Fig. 12 shows that LEF-YOLO has high detection confidence and fewer problems with overlapping detection frames (Fig. 12f).

Comparison of detection speed between our proposed algorithm and state-of-the-art algorithms

In terms of model detection speed, LEF-YOLO has several advantages, including computational efficiency, fast detection speed, and a small model size. As shown in Table 3, the parameters of LEF-YOLO are the smallest when compared to the other models, with a size of only 1.33 compared to the basic model YOLOv5, YOLOv5 has a smaller FPS, but the proposed algorithm's model size and FLOPs (floating-point operations per second) are smaller, with a model size of 3.11 M and FLOPs of 2.7 G, while still achieving a processing speed of 61 FPS. This meets the purpose of real-time detection. The model has a smaller size compared to other models. In contrast, the model introduces multiple depthwise separable convolutions, which allows for a reduction in parameters and computation. The operation of DW convolution and PW convolution is also simpler, and the computation of forward and backward propagation can be performed faster compared to the traditional convolutional layers. However, the model has fewer channel dimensions and network layers, making its structure not as complex as other models. Specifically, the model has smaller channel dimensions than YOLOv5s-ShuffleNetv2 and YOLOv5s-MobileNetv3l and fewer network layers than YOLOv5s-EfficientNet. Hence LEF-YOLO is lightweight and can be carried on mobile devices, and its detection speed is fast enough to meet the requirements for real-time detection of extreme wildfire targets. It is superior to other algorithms, demonstrating its effectiveness for real time, intelligent, and accurate detection of extreme wildfires.

Comparison of our proposed algorithm with lightweight algorithms

We compared LEF-YOLO with other lightweight models, including YOLOv5s-ShuffleNetv2, YOLOv5s-EfficientNet,

and YOLOv5s-MobileNetv3, which were also improved on YOLOv5s. The P-R curves of LEF-YOLO and other lightweight algorithms are in Fig. 13. The area enclosed by the P-R curve of LEF-YOLO is larger, indicating higher detection accuracy. LEF-YOLO has higher precision, recall, and mAP than these models (Fig. 14, Table 4). LEF-YOLO uses the bottleneck structure of MobileNetv3 and the channel shuffle structure of ShuffleNetv2 in its architecture. This greatly reduces the model's computation time and size. YOLOv5s-MobileNetv3 has the smallest computation time and size among them, and our LEF-YOLO is very close to it, ensuring high accuracy while being sufficiently lightweight. This makes LEF-YOLO suitable for the real time detection of extreme wildfires in complex scenarios.

Ablation experiments

We applied the new backbone structure, CA module and SPPF structure in LEF-YOLO. To verify the validity of the model, we performed ablation experiments. As shown in Table 5, compared with M1, we found that after applying the new backbone structure, the parameters of M2 are reduced by 5.37 M, the inference time at CPU is reduced by 17 ms, and the FLOPs are only 2.7 In addition, the mAP is only reduced by 4.9%. This verifies that the new backbone can greatly reduce the model's computation with higher accuracy and give the model a speedup. When comparing M1 and M3, after replacing SPPF, the mAP of M3 is improved by 0.9%, and the inference time in CPU is reduced by 4.1 ms. This shows that SPPF can bring performance improvements in accuracy and speed to the model without introducing computational effort. By comparing M1 and M4, we find that after introducing the CA module, the FLOPs increase by only 0.5 G, the parameters increase by only 0.06 M, but the mAP increases by 1.7%. This verifies that the CA module can improve the accuracy of the model at a lower computational volume. Finally, training M5, which combines all methods, achieves the best results. Compared with M1, M5 has a 4.6% increase in mAP, a 5.37 m decrease in parameters, a 13.1 G decrease in FLOPs, and a 15.8 ms increase in CPU inference speed. These ablation experiments show that after the improvement of the three methods, the model absorbs the advantages of each method, and the model brings a significant improvement in the average detection speed and accuracy. Thus the model meets the performance requirements for detecting extreme wildfires.

Visualisation of the results of the proposed algorithm

Neural network-based deep learning is a type of machine learning that is known for being less interpretable and is often regarded as a 'black box' approach. Although the previous section has demonstrated the excellent performance of the proposed model, it is difficult to determine whether deep learning primarily focuses on extreme wildfire

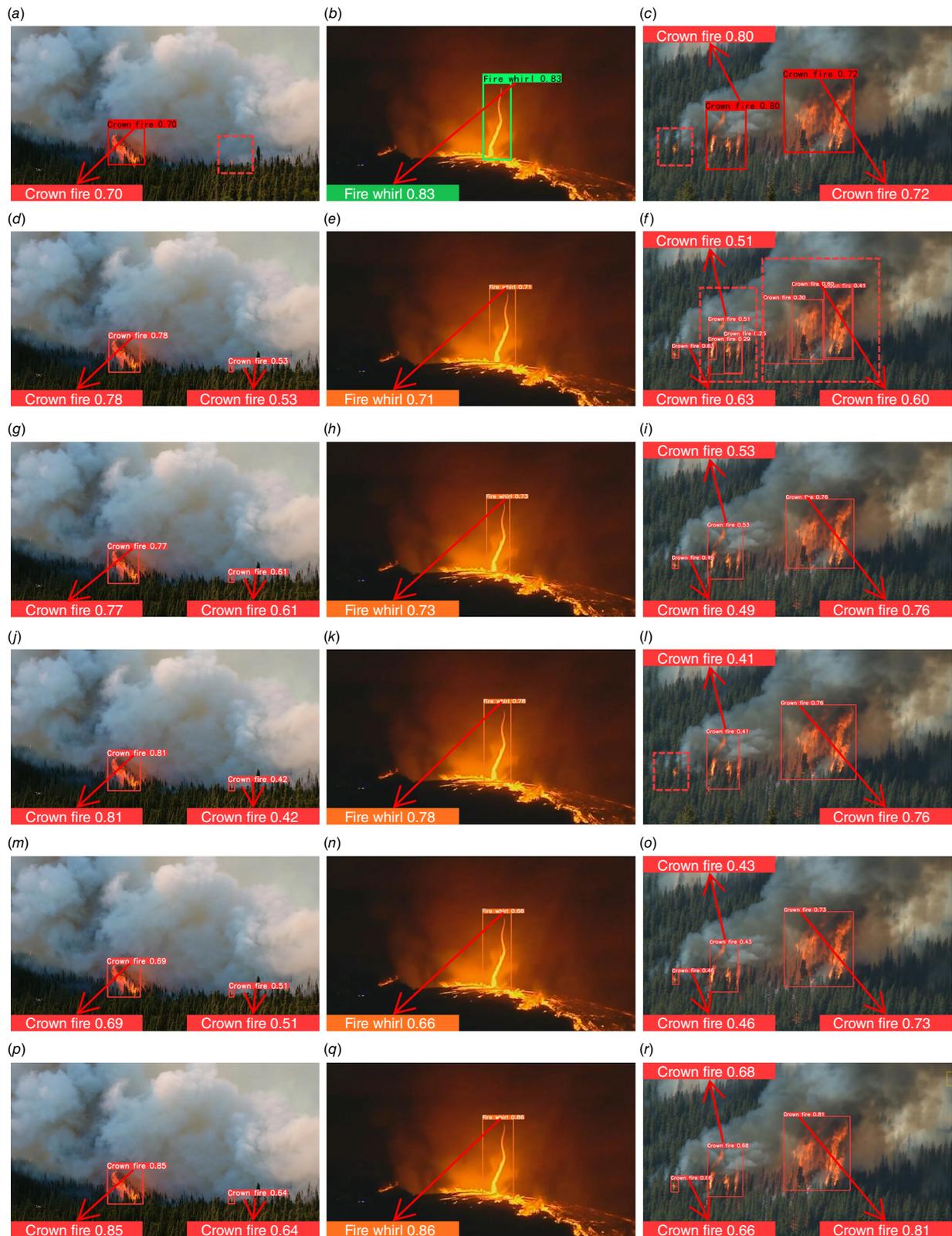
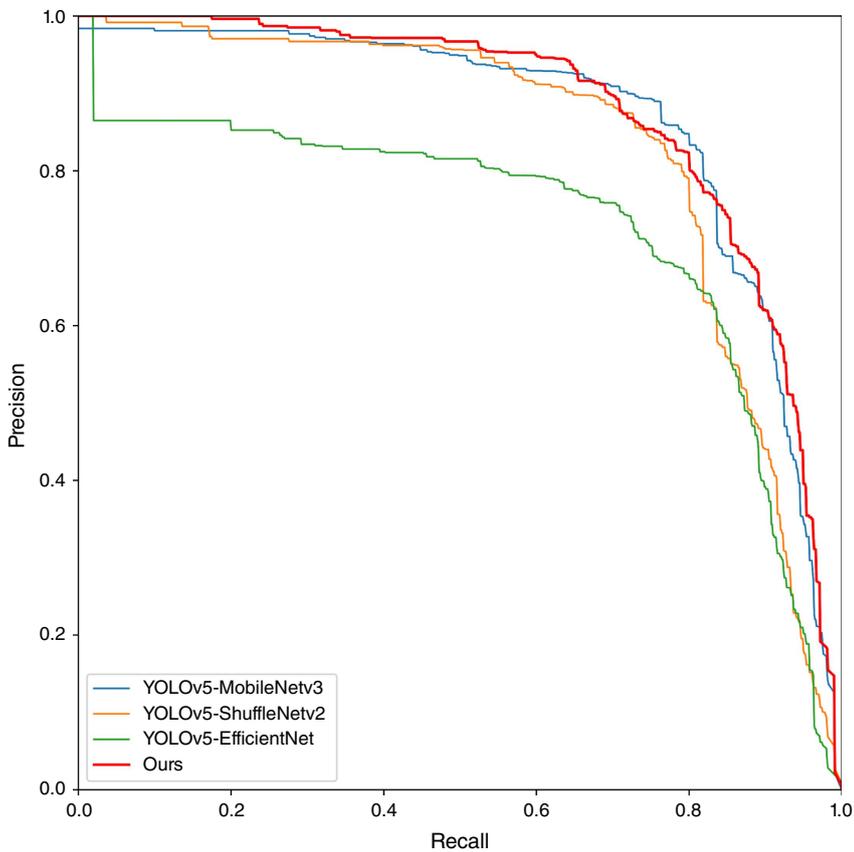


Fig. 12. Comparison of detection results of different algorithms. (a–c) SSD-300, (d–f) YOLOv3, (g–i) YOLOv4s-mish, (j–l) YOLOv5s, (m–o) M-YOLOv5s, and (p–r) LEF-YOLO (this paper).

Table 3. Performance comparison between different algorithms.

Algorithms	Precision (%)	Recall (%)	mAP (%)	Parameters (M)	Model size (M)	FLOPs (G)	Processing speed (FPS)	Inference time in CPU (ms)
SSD-300	72.6	50.8	63.1	24.5	92.13	62.8	32	163.2
YOLOv3	69.0	66.5	68.6	58.66	117.73	154.6	17	284.5
YOLOv4s-mish	70.4	77.5	80.0	8.69	17.73	20.6	67	157.3
YOLOv5s	82.7	78.0	83.3	6.70	13.75	15.8	75	87.6
Ours	82.1	83.6	87.9	1.33	3.11	2.7	61	71.8

**Fig. 13.** P-R curves between LEF-YOLO and different lightweight algorithms.

features or noisy features since feature extraction in deep learning relies on hidden layer networks. To address this issue, we employed the Grad-CAM algorithm (Selvaraju *et al.* 2017) to visualise the results of our proposed algorithm. This algorithm allows for a visual interpretation of the convolutional network and provides insights into how the network identifies extreme wildfire features over noise features. As shown in Fig. 15, the results indicate that LEF-YOLO can effectively focus on extreme wildfire features and precisely locate them within the images. We also visualised the feature maps of the first CA output layer (Fig. 16), which extracted the key features of the extreme wildfires. The feature maps show that the crown fire is conical and ray-shaped, the firelines merge in jagged and scissor shapes, the

fire whirl is conical and cylindrical, and the spot fire is scattered, which also matches the description of the flame shapes in Table 1. This can be attributed to the unique network structure of LEF-YOLO, which incorporates a CA and employs multiscale fusion methods in feature extraction. Thus, the feature extraction capability of LEF-YOLO is supported by the visual analysis of the results.

Conclusion

This article presents an extreme wildfire detection algorithm termed LEF-YOLO, based on lightweight deep learning. The LEF-YOLO model is an improvement of YOLOv5. The model

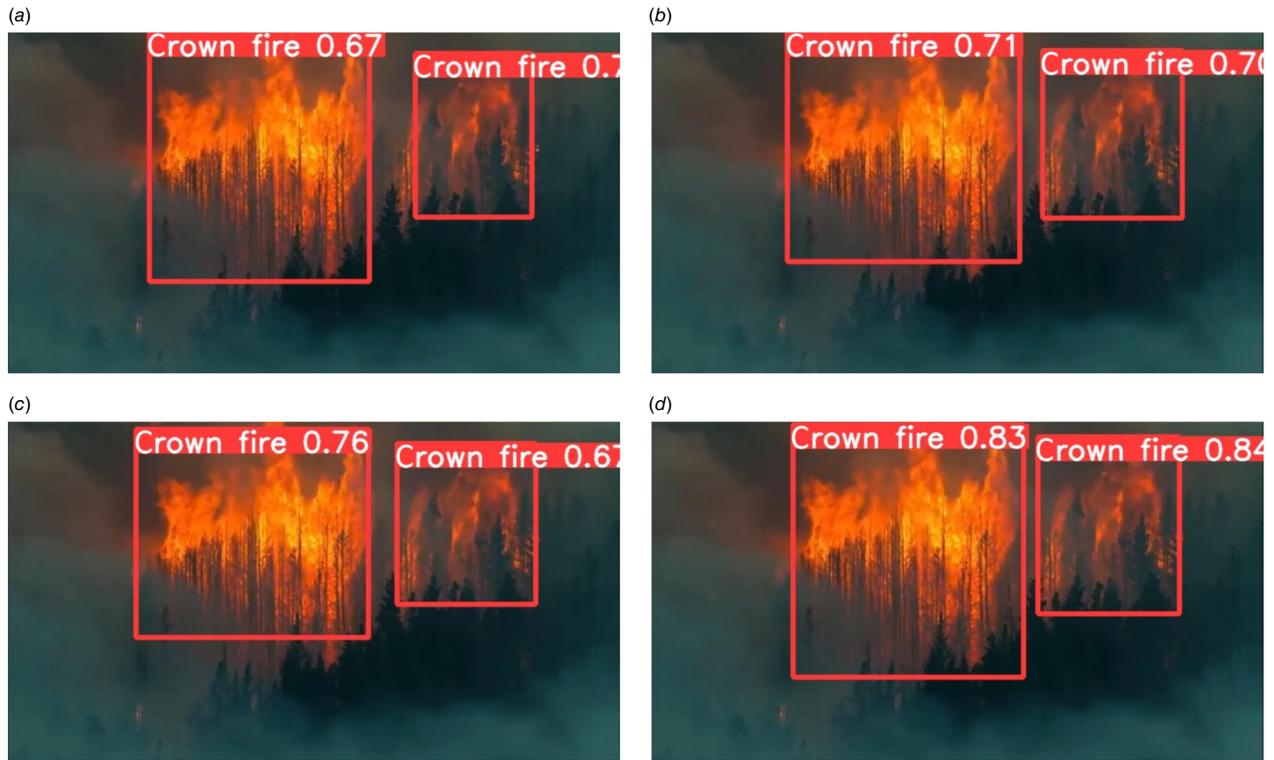


Fig. 14. Comparison of detection results of different lightweight algorithms. (a) YOLOv5-ShuffleNet2, (b) YOLOv5-EfficientNet, (c) YOLOv5-MobileNet3, and (d) LEF-YOLO (this paper).

Table 4. Performance comparison between different lightweight algorithms.

Algorithms	Precision (%)	Recall (%)	mAP (%)	Parameters (M)	Model size (M)	FLOPs (G)	Processing speed (FPS)	Inference time in CPU (ms)
YOLOv5s-ShuffleNet2	81.5	79.1	83.4	3.04	6.39	5.9	71	74.1
YOLOv5s-MobileNet3l	81.6	82.4	86.7	2.51	5.39	4.8	78	129.7
YOLOv5s-EfficientNet	67.9	72.6	72.8	5.95	12.38	9.8	54	120.7
Ours	82.1	83.6	87.9	1.33	2.99	2.7	61	71.8

Table 5. Improved algorithm for ablation study.

Model	Method	mAP (%)	Parameters (M)	FLOPs (G)	Inference time in CPU (ms)
M1	YOLOv5s	83.3	6.70	15.8	87.6
M2	M1 + (LEF-YOLO-Backbone)	78.4	1.33	2.7	70.6
M3	M1 + SPPF	84.2	6.70	15.8	83.5
M4	M1 + CA	85.0	6.76	16.3	97.3
M5 (ours)	M2 + SPPF + CA	87.9	1.33	2.7	71.8

introduces the bottleneck structure from Mobilenetv3 and replaces traditional convolution with depthwise separable convolution to reduce the number of parameters of the model. After convolution with 3×3 and 5×5 kernels, the corresponding feature maps are fused with multiscale

feature fusion. At the same time, the model incorporates the CA module and SPPF structure to improve the feature extraction ability for extreme wildfire feature maps thereby ensuring the accuracy of the model detection. Comparative experiments show that LEF-YOLO has a higher model size

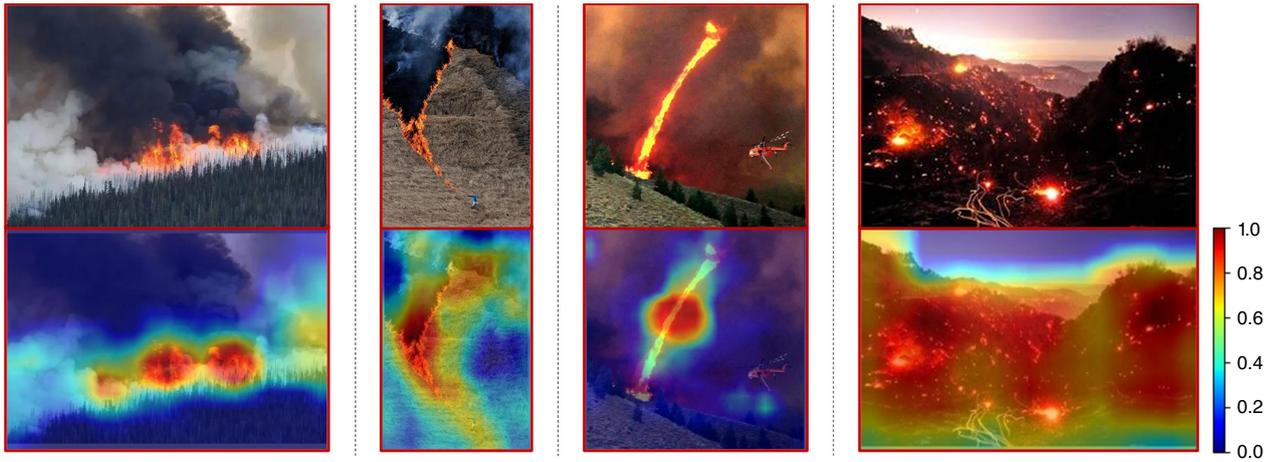


Fig. 15. Visualisation of partial model detection results. The colour-coded layers represent areas of attention. The closer the plotted area is to 1 (red) means that the plotted target is more attended to, and the closer it is to 0 (blue-violet) the less attended to.

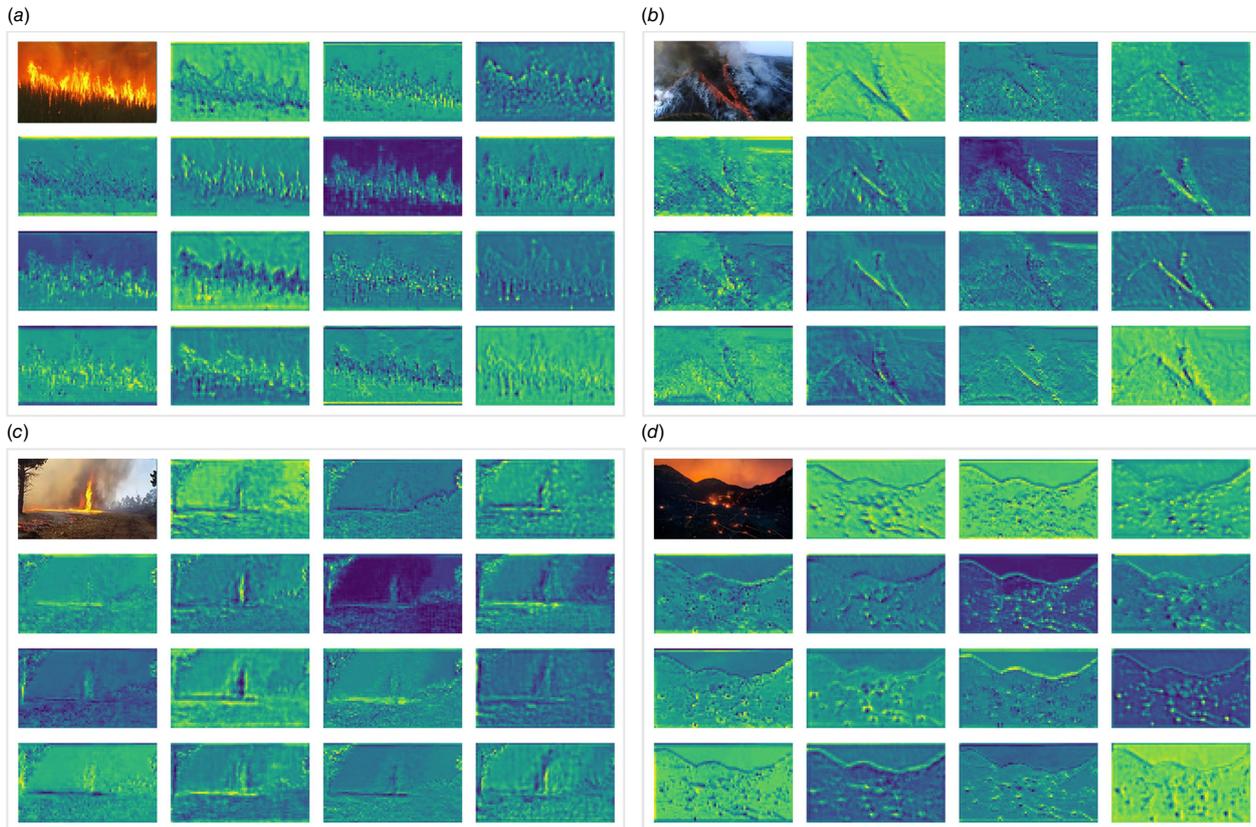


Fig. 16. The feature maps of the four extreme wildfires after depthwise separable convolutional fusion and Coordinate Attention. The first one is the original map, and the rest are partial feature maps. (a) Crown fire, (b) firelines merging, (c) fire whirl, and (d) spot fire.

and computational complexity than YOLOv5 but with a model size and FLOPs of only 3.11 M and 2.7 G, respectively. Compared to mainstream models, it has higher detection accuracy, with an mAP of 87.9%. Thus LEF-YOLO can

ensure real-time detection of extreme wildfires in practical scenarios.

In future work, more extreme wildfire types and images of the various stages of extreme wildfire evolution will be

added to the dataset to train a model with better detection performance. In addition, we will improve our detection method to overcome the limitation of detecting extreme wildfires with multi-stage development and further improve the ability of our model to detect spot fires. And we plan to use the latest YOLO or other models to build even more capable models to further improve the overall performance of detecting each type of extreme wildfire, while ensuring that these models are lightweight enough to be more easily deployed on personal portable devices or firefighting robots.

References

- Almeida JS, Huang C, Nogueira FG, Bhatia S, De Albuquerque VHC (2022) EdgeFireSmoke: A Novel Lightweight CNN Model for Real-Time Video Fire-Smoke Detection. *IEEE Transactions on Industrial Informatics* 18(11), 7889–7898. doi:10.1109/Tii.2021.3138752
- Azim MR, Keskin M, Do N, Gül M (2022) Automated classification of fuel types using roadside images via deep learning. *International Journal of Wildland Fire* 31(10), 982–987. doi:10.1071/Wf21136
- Bao W, Yang X, Liang D, Hu G, Yang X (2021) Lightweight convolutional neural network model for field wheat ear disease identification. *Computers and Electronics in Agriculture* 189(4), 106367. doi:10.1016/j.compag.2021.106367
- Barpoutis P, Dimitropoulos K, Kaza K, Grammalidis N (2019) Fire Detection from Images Using Faster R-CNN and Multidimensional Texture Analysis. In 'ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)', 12–17 May 2019, Brighton, UK. pp. 8301–8305. doi:10.1109/ICASSP.2019.8682647
- Castro RF, Morgan P, Fernandes P, Hoffman C (2021) Extreme Fires. In 'Fire Science'. pp. 175–257. (Springer: Cham) doi:10.1007/978-3-030-69815-7_8
- Chen A (2022) Evaluating the relationships between wildfires and drought using machine learning. *International Journal of Wildland Fire* 31(3), 230–239. doi:10.1071/Wf21145
- Chen XW, Hopkins B, Wang H, O'Neill L, Afghah F, Razi A, Fule P, Coen J, Rowell E, Watts A (2022) Wildland Fire Detection and Monitoring Using a Drone-Collected RGB/IR Image Dataset. *IEEE Access* 10, 121301–121317. doi:10.1109/Access.2022.3222805
- Colston K, Flik R (2012) 'Forest management for resilience, adaptation and watershed protection.' (Nova Science Publishing: USA)
- Cui MD, Lou YY, Ge YL, Wang KQ (2023) LES-YOLO: A lightweight pinecone detection algorithm based on improved YOLOv4-Tiny network. *Computers and Electronics in Agriculture* 205, 107613. doi:10.1016/j.compag.2023.107613
- Dai Y, Liu W, Wang H, Xie W, Long K (2022) YOLO-Former: Marrying YOLO and Transformer for Foreign Object Detection. *IEEE Transactions on Instrumentation and Measurement* 71, 1–14. doi:10.1109/TIM.2022.3219468
- Dang FY, Chen D, Lu YZ, Li ZJ (2023) YOLOWeeds: A novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems. *Computers and Electronics in Agriculture* 205, 107655. doi:10.1016/j.compag.2023.107655
- Department of Forestry and Fire Management (2013) Yarnell Hill Fire Report Now Available. Available at <https://dffm.az.gov/yarnell-hill-report-available> [verified 23 September 2013]
- Diao ZH, Yan JN, He ZD, Zhao SN, Guo PL (2022) Corn seedling recognition algorithm based on hyperspectral image and lightweight-3D-CNN. *Computers and Electronics in Agriculture* 201, 107343. doi:10.1016/j.compag.2022.107343
- Gómez-Vázquez I, Fernandes PM, Arias-Rodil M, Barrio-Anta M, Castedo-Dorado F (2014) Using density management diagrams to assess crown fire potential in Pinus pinaster Ait. stands. *Annals of Forest Science* 71(4), 473–484. doi:10.1007/s13595-013-0350-4
- Gonzalez-Huitron V, León-Borges JA, Rodríguez-Mata AE, Amabilis-Sosa LE, Ramírez-Pereda B, Rodríguez H (2021) Disease detection in tomato leaves via CNN with lightweight architectures implemented in Raspberry Pi 4. *Computers and Electronics in Agriculture* 181, 105951. doi:10.1016/j.compag.2020.105951
- Hantson S, Andela N, Goulsten ML, Randerson JT (2022) Human-ignited fires result in more extreme fire behavior and ecosystem impacts. *Nature Communications* 13(1), 2717. doi:10.1038/s41467-022-30030-2
- Hou Q, Zhou D, Feng J (2021) Coordinate Attention for Efficient Mobile Network Design. In '2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)', Nashville, TN, USA, 20–25 June 2021. pp. 13708–13717. doi:10.1109/CVPR46437.2021.01350
- Howard A, Sandler M, Chu G, Chen LC, Chen B, Tan MX, Wang WJ, Zhu YK, Pang RM, Vasudevan V, Le QV, Adam H (2019) Searching for MobileNetV3. In '2019 IEEE/CVF International Conference on Computer Vision (ICCV)', Seoul, Korea, 27 October–2 November 2019. pp. 1314–1324. doi:10.1109/ICCV.2019.00140
- Hsu WY, Lin WY (2021) Ratio-and-Scale-Aware YOLO for Pedestrian Detection. *IEEE Transactions on Image Processing* 30, 934–947. doi:10.1109/TIP.2020.3039574
- Jocher G (2020) yolov5. Code repository. Available at <https://github.com/ultralytics/yolov5>
- Lei J, Deng WY, Mao SH, Tao Y, Wu HG, Xie CG (2023) Flame geometric characteristics of large-scale pool fires under controlled wind conditions. *Proceedings of the Combustion Institute* 39(3), 4021–4029. doi:10.1016/j.proci.2022.07.132
- Li JW, Li XW, Chen CC, Zheng HR, Liu NY (2018) Three-dimensional dynamic simulation system for forest surface fire spreading prediction. *International Journal of Pattern Recognition and Artificial Intelligence* 32(8), 1850026. doi:10.1142/S021800141850026X
- Lin TY, Dollar P, Girshick R, He KM, Hariharan B, Belongie S (2017) Feature Pyramid Networks for Object Detection. In '2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)', Honolulu, HI, USA, 21–26 July 2017. pp. 936–944. doi:10.1109/CVPR.2017.106
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD: Single shot multibox detector. In 'Computer Vision - ECCV 2016. ECCV Lecture Notes in Computer Science'. (Eds B Leibe, J Matas, N Sebe, M Welling) pp. 21–37. (Springer: Cham) doi:10.1007/978-3-319-46448-0_2
- Liu NA, Lei J, Gao W, Chen HX, Xie XD (2021) Combustion dynamics of large-scale wildfires. *Proceedings of the Combustion Institute* 38(1), 157–198. doi:10.1016/j.proci.2020.11.006
- Ma N, Zhang X, Zheng HT, Sun J (2018) Shufflenet v2: Practical guidelines for efficient CNN architecture design. In 'Computer Vision - ECCV 2018. ECCV 2018 Lecture Notes in Computer Science'. pp. 116–131. (Springer: Cham) doi:10.1007/978-3-030-01264-9_8
- Muhammad K, Ahmad J, Lv Z, Bellavista P, Yang P, Baik SW (2019) Efficient Deep CNN-Based Fire Detection and Localization in Video Surveillance Applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 49(7), 1419–1434. doi:10.1109/Tsmc.2018.2830099
- National Wildfire Coordinating Group (2020) Extreme Fire Behavior. Available at <https://www.nwcg.gov/term/glossary/extreme-fire-behavior> [verified 2020]
- Nguyen DT, Nguyen TN, Kim H, Lee HJ (2019) A High-Throughput and Power-Efficient FPGA Implementation of YOLO CNN for Object Detection. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 27(8), 1861–1873. doi:10.1109/TVLSI.2019.2905242
- Pritam D, Dewan JH (2017) Detection of fire using image processing techniques with LUV color space. In '2017 2nd International Conference for Convergence in Technology (I2CT)', Mumbai, India, 7–9 April 2017. pp. 1158–1162. doi:10.1109/I2CT.2017.8226309
- Qiang XH, Zhou GX, Chen AB, Zhang X, Zhang WZ (2021) Forest fire smoke detection under complex backgrounds using TRPCA and TSVB. *International Journal of Wildland Fire* 30(5), 329–350. doi:10.1071/Wf20086
- Qin L, Shi Y, He YH, Zhang JR, Zhang XS, Li YJ, Deng T, Yan HM (2022) ID-YOLO: Real-Time Salient Object Detection Based on the Driver's Fixation Region. *IEEE Transactions on Intelligent Transportation Systems* 23(9), 15898–15908. doi:10.1109/Tits.2022.3146271
- Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In '2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)', Las Vegas, NV, USA, 27–30 June 2016. pp. 779–788. doi:10.1109/CVPR.2016.91
- Sadykova D, Pernebayeva D, Bagheri M, James A (2020) IN-YOLO: Real-Time Detection of Outdoor High Voltage Insulators Using UAV

- Imaging. *IEEE Transactions on Power Delivery* 35(3), 1599–1601. doi:10.1109/TPWRD.2019.2944741
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In '2017 IEEE International Conference on Computer Vision (ICCV)', Venice, Italy, 22–29 October 2017. pp. 618–626. doi:10.1109/ICCV.2017.74
- Sharma LK, Gupta R, Fatima N (2022) Assessing the predictive efficacy of six machine learning algorithms for the susceptibility of Indian forests to fire. *International Journal of Wildland Fire* 31(8), 735–758. doi:10.1071/Wf22016
- Shees A, Ansari MS, Varshney A, Asghar MN, Kanwal N (2023) FireNet-v2: Improved Lightweight Fire Detection Model for Real-Time IoT Applications. *Procedia Computer Science* 218, 2233–2242. doi:10.1016/j.procs.2023.01.199
- Sudhakar S, Vijayakumar V, Sathiy Kumar C, Priya V, Ravi L, Subramaniaswamy V (2020) Unmanned Aerial Vehicle (UAV) based Forest Fire Detection and monitoring for reducing false alarms in forest-fires. *Computer Communications* 149, 1–16. doi:10.1016/j.comcom.2019.10.007
- Tedim F, Leone V, Amraoui M, Bouillon C, Coughlan MR, Delogu GM, Fernandes PM, Ferreira C, McCaffrey S, McGee TK, Parente J, Paton D, Pereira MG, Ribeiro LM, Viegas DX, Xanthopoulos G (2018) Defining Extreme Wildfire Events: Difficulties, Challenges, and Impacts. *Fire* 1(1), 9. doi:10.3390/fire1010009
- The People's Government of Sichuan Province (2020) "3.30" Forest Fire Event Investigation Result in Xichang City, Liangshan Prefecture. Available at <https://www.sc.gov.cn/10462/10464/13722/2020/12/21/76441b52bf034463946b09f61876d3f9.shtml> [verified 21 December 2020]
- Tohidi A, Gollner MJ, Xiao HH (2018) Fire Whirls. *Annual Review of Fluid Mechanics* 50, 187–213. doi:10.1146/annurev-fluid-122316-045209
- Viegas DX, Simeoni A (2011) Eruptive Behaviour of Forest Fires. *Fire Technology* 47, 303–320. doi:10.1007/s10694-010-0193-6
- Wang T, Shi L, Yuan P, Bu LP, Hou XG (2017) A New Fire Detection Method Based on Flame Color Dispersion and Similarity in Consecutive Frames. In '2017 Chinese Automation Congress (CAC)', Jinan, China, 20–22 October 2017. pp. 151–156. doi:10.1109/CAC.2017.8242754
- Wang C-Y, Liao H-YM, Wu Y-H, Chen P-Y, Hsieh J-W, Yeh I-H (2020) CSPNet: A new backbone that can enhance learning capability of CNN. In '2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)', Seattle, WA, USA, 14–19 June 2020. pp. 1571–1580. doi:10.1109/CVPRW50498.2020.00203
- Werth PA, Potter BE, Clements CB, Finney MA, Goodrick SL, Alexander ME, Cruz MG, Forthofer JA, McAllister SS (2011) Synthesis of knowledge of extreme fire behavior: Volume I for fire managers. General Technical Report PNW-GTR-54. (USDA Forest Service, Pacific Northwest Research Station: Portland, OR)
- Werth PA, Potter BE, Alexander ME, Clements CB, Cruz MG, Finney MA, Forthofer JM, Goodrick SL, Hoffman C, Jolly WM, McAllister SS, Ottmar RD, Parsons RA (2016) Synthesis of Knowledge of Extreme Fire Behavior: Volume 2 for fire behavior specialists, researchers, and meteorologists. General Technical Report PNW-GTR-891. (USDA Forest Service, Pacific Northwest Research Station: Portland, OR)
- Wu XQ, Tan GH, Zhu NB, Chen ZL, Yang Y, Wen HX, Li KL (2021) CacheTrack-YOLO: Real-Time Detection and Tracking for Thyroid Nodules and Surrounding Tissues in Ultrasound Videos. *Ieee Journal of Biomedical and Health Informatics* 25(10), 3812–3823. doi:10.1109/Jbhi.2021.3084962
- Wu YF, Li JW, Bi S, Zhu X, Wang QF (2023) Research on Improved Ant Colony Algorithm for Mountain Hiking Emergency Rescue Path Planning. *Journal of Geo-Information Science* 25(1), 90–101. doi:10.12082/dqxxkx.2023.220535
- Zaidi SSA, Ansari MS, Aslam A, Kanwal N, Asghar M, Lee B (2022) A survey of modern deep learning based object detection models. *Digital Signal Processing* 126, 103514. doi:10.1016/j.dsp.2022.103514

Data availability. The data supporting this study will be shared on reasonable request sent to the corresponding author.

Conflicts of interest. The authors declare no conflict of interest.

Declaration of funding. This research was supported by the National Key Research and Development Program of China (Grant No. 2022YFC3003000), National Natural Science Foundation of China (Grant No. 32071776), Natural Science Foundation of Fujian Province, China (Grant No. 2020J01465), and China Postdoctoral Science Foundation (Grant No. 2018M640597).

Author affiliations

^ACollege of Physics and Information Engineering, Fuzhou University, Fuzhou, 350116, China.

^BCollege of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin, China.

^CZijin School of Geology and Mining, Fuzhou University, Fuzhou 350116, China.

^DHuadong Engineering Corporation Limited, Hangzhou, 311122, China.