remote sensing

MDPI

Article

LUFFD-YOLO: A Lightweight Model for UAV Remote Sensing Forest Fire Detection Based on Attention Mechanism and Multi-Level Feature Fusion

Yuhang Han [1], Bingchen Duan [1], Renxiang Guan [2], Guang Yang [3] and Zhen Zhen [3,*]

[1] College of Aulin, Northeast Forestry University, Harbin 150040, China; hanyh@nefu.edu.cn (Y.H.); 505586225@nefu.edu.cn (B.D.)
[2] College of Computer, National University of Defense Technology, Changsha 410073, China; renxiangguan@nudt.edu.cn
[3] Key Laboratory of Sustainable Forest Ecosystem Management-Ministry of Education, School of Forestry, Northeast Forestry University, Harbin 150040, China; yangguang@nefu.edu.cn
* Correspondence: zhenzhen@nefu.edu.cn; Tel.: +86-0451-8219-1215

Abstract: The timely and precise detection of forest fires is critical for halting the spread of wildfires and minimizing ecological and economic damage. However, the large variation in target size and the complexity of the background in UAV remote sensing images increase the difficulty of real-time forest fire detection. To address this challenge, this study proposes a lightweight YOLO model for UAV remote sensing forest fire detection (LUFFD-YOLO) based on attention mechanism and multi-level feature fusion techniques: (1) GhostNetV2 was employed to enhance the conventional convolution in YOLOv8n for decreasing the number of parameters in the model; (2) a plug-and-play enhanced small-object forest fire detection C2f (ESDC2f) structure was proposed to enhance the detection capability for small forest fires; (3) an innovative hierarchical feature-integrated C2f (HFIC2f) structure was proposed to improve the model's ability to extract information from complex backgrounds and the capability of feature fusion. The LUFFD-YOLO model surpasses the YOLOv8n, achieving a 5.1% enhancement in mAP and a 13% reduction in parameter count and obtaining desirable generalization on different datasets, indicating a good balance between high accuracy and model efficiency. This work would provide significant technical support for real-time forest fire detection using UAV remote-sensing images.

Keywords: attention mechanism; feature fusion; forest fire; lightweight network; UAV remote sensing images

check for updates

Citation: Han, Y.; Duan, B.; Guan, R.; Yang, G.; Zhen, Z. LUFFD-YOLO: A Lightweight Model for UAV Remote Sensing Forest Fire Detection Based on Attention Mechanism and Multi-Level Feature Fusion. *Remote Sens.* **2024**, *16*, 2177. https://doi.org/10.3390/rs16122177

Academic Editors: Luis A. Ruiz, António Bento-Gonçalves, Ting Yun and Fuquan Zhang

Received: 23 April 2024
Revised: 8 June 2024
Accepted: 13 June 2024
Published: 15 June 2024

1. Introduction

Forest fires are a type of natural disaster with high frequency and devastating power [1–3], and they profoundly impact human life, socio-economics, and natural ecosystems worldwide [4]. Forest fires consume vast forest resources, lead to biodiversity loss, and generate significant greenhouse gas emissions, accelerating global climate change [5,6]. Moreover, population growth, urbanization, and certain human activities [7], such as illegal logging and changes in land cover, significantly increase the risk and intensity of forest fires. Thus, the timely and precise detection of forest fires is paramount for the protection of forest ecosystems.

Given that forest fires spread rapidly in environments rich in oxygen [8] and with swift air currents [9], prompt detection becomes critically important. Traditional forest fire detection primarily relies on manual patrols [10] and remote sensing technology [11,12]. While manual patrols can directly assess fire situations, this method is resource-intensive, time-consuming, covers limited areas, and carries high risks. The advancement of remote sensing technology offers convenience for efficient and rapid forest fire detection. Thermal infrared sensors are commonly applied to detect forest fires by capturing variations in the intensity of infrared radiation emitted from high-temperature areas (such as fire sources) in forest regions [13]. Satellite images with high temporal resolution, such as Himawari-9 [14] and MODIS [15], enable continuous and extensive monitoring of forest fires over a wide area. However, the limited spatial resolution of them significantly impairs the ability to detect small forest fires. The utilization of Unmanned Aerial Vehicle (UAV) remote sensing technology [16] enables the capture of images with much higher spatial resolution. Moreover, its versatility and agility make it highly effective in complex landscapes. Despite the use of advanced UAV remote sensing technologies, human participation remains essential in determining the presence of forest fires. This process also requires a substantial amount of verification effort. Hence, to tackle these challenges, real-time fire detection methods employing computer vision and UAV technology have been implemented, enhancing the accuracy and efficiency of forest fire detection [17].

In the initial stages, traditional machine learning methods relied extensively on feature engineering to discern characteristics linked to forest fires. These features were then paired with appropriate machine-learning models to enable the detection of forest fire occurrences. For example, Yang et al. [18] introduced an enhanced support vector machine model, PreVM, by enhancing the approach to L1 regularization. This effectively resolves the problem of non-paired instances of forest fire characteristic samples. Maeda et al. [19] utilized environmental characteristics and the sun zenith angle as features and a random forest classifier to identify forest fires, resulting in a notable improvement in the accuracy of fire identification. Nevertheless, conventional approaches primarily detect feature sets by means of manual selection and feature design, significantly depending on expert knowledge, and the process of selecting features can be time-consuming. Moreover, it cannot guarantee accurate identification of the optimal feature sets. Therefore, deep learning algorithms [20,21] have a distinct advantage in automatically extracting elaborate feature representations from raw data [22].

Deep learning [23–27] is a machine learning technology that uses artificial neural networks to mimic the cognitive processes of the human brain. It is used to analyze data, recognize patterns, and make judgments [28]. The ongoing advancement of deep learning has made the identification of forest fires in many areas a central focus of research. From a computer vision perspective, forest fire detection using deep learning techniques mainly falls into two categories: two-stage detection methods [29] and one-stage detection methods [30–32]. In the context of forest fire detection, the two-stage object detection method consists of two phases: the initial phase involves identifying potential areas where fires may be present, and the following phase focuses on improving the accuracy of fire classification and adjusting the bounding box regression for these areas (e.g., [33,34]). Nevertheless, the major disadvantage of the two-stage object detection method is its reliance on significant computer resources. The limited speed of both training and inference hinders the ability to identify forest fires in real-time. To address this issue, the YOLO series [35] of one-stage object detection models are commonly employed. The distinctive feature lies in employing anchor boxes for predicting the position and category of targets. YOLO series models exhibit faster detection speeds in comparison to two-stage models. They employ direct target detection in photos, bypassing the need to generate candidate zones, making them extensively utilized in forest fire detection. Attention mechanisms [36], which draw inspiration from human biological systems, enable neural networks to selectively focus on pertinent information when handling vast quantities of input. As a result, they are widely used in combination with YOLO neural networks. For example, Luo et al. [37] proposed a YOLOX algorithm that integrates the Swin Transformer architecture, convolutional block attention module, and slim neck structure for forest fire detection. This algorithm demonstrates superior detection precision and enhanced capability for accurately recognizing locations even in complex conditions, achieving an *mAP* value of 92.26%. Nevertheless, it is notable for its extensive parameterization and heightened demand for computational resources, which hampers its suitability for real-time forest fire detection. Xue et al. [38] addressed the issue of low recognition accuracy for complex background forest fires, proposed an FCDM model based on YOLOv5. It improved the detection capability for different types of fires by modifying the bounding box loss function and introducing the convolutional block attention mechanism, with ground fire detection rates of 83.1% and crown fire detection rates of 90.6%. Chen et al. [39] introduced an LMDFS model based on YOLOv7. They improved the smoke feature extraction capabilities by incorporating a coordinate attention mechanism and a content-aware reassembly feature upsampling technique. The accuracy showed a 5.9% improvement in comparison to the baseline model (YOLOv7). However, it still has limitations when faced with irregular forest fire images or forest fire problems that involve tiny targets (i.e., small forest fires).

Despite significant advancements in algorithm development, the use of UAV images for forest fire detection continues to encounter many obstacles: (1) The motion of UAVs in various mountainous and forested settings amplifies the intricacy of the background in the images. This intricacy is intensified by elements such as vegetation, unpredictable meteorological circumstances, variations in illumination, and the existence of clouds, fog, and smoke; (2) It is crucial to use a network with a more complex structure that can extract an optimal number of features. However, this is a notable obstacle in terms of detecting forest fires in real time. Thus, to tackle these difficulties, this study proposes a lightweight UAV remote-sensing forest fire detection algorithm, called LUFFD-YOLO. This model is primarily designed to detect UAV-based images with red, green, and blue spectral bands. It successfully achieved a trade-off between model efficiency, gained by minimizing the number of parameters, and retained high accuracy in detecting forest fires. The main contributions of this paper are as follows:

(1) Innovatively, the LUFFD-YOLO model adopts the GhostNetV2 structure to optimize the conventional convolutions of the YOLOv8n backbone layer, resulting in a more efficient and streamlined network design. This significantly reduces the model's complexity and computational requirements.

(2) This study proposes a plug-and-play enhanced small-object forest fire detection C2f (ESDC2f) module that utilizes the Multi-Head Self Attention (MHSA) mechanism to boost the detection capability for small objects and compensate for the loss caused by lightweight in LUFFD-YOLO model. It greatly enhances the capability to extract features from various subspaces of UAV images, hence increasing the accuracy of forest fire detection.

(3) A hierarchical feature-integrated C2f (HFIC2f) model, using the SegNeXt attention mechanism, has been proposed to effectively tackle the problem of low accuracy in detecting forest fire objects against complicated backgrounds.

The following sections of this paper are organized as follows: Section 2 outlines the datasets and introduces both the YOLOv8 model (as a baseline) and the LUFFD-YOLO model. Sections 3 and 4 detail the experimental design and case analysis, respectively. Finally, Section 5 provides conclusions and highlights future research avenues.

2. Materials and Methods

2.1. Datasets

This study performs a comprehensive evaluation of the LUFFD-YOLO model on a large public forest fire UAV remote sensing dataset (M4SFWD), a comprehensive remote sensing dataset for large-scale forest fires (FLAME), and a manually constructed UAV-based forest fire dataset (SURSFF). The images in the dataset are captured with the red, green, and blue spectral bands.

The M4SFWD dataset [40] is a large-scale forest fire remote sensing dataset captured from drone perspectives. This dataset includes a range of terrain types, meteorological circumstances, light intensities, and varying quantities of forest fire occurrences. This dataset includes 3974 images and its label categories include fire and smoke. The collection contains images of various sizes, including 1480 × 684 and 1280 × 720, to cater to different scenarios. Some images in the collection have excessively large aspect ratios, making them unsuitable for direct use in training. Therefore, during the model training and testing processes, all images were resized to a uniform size of 640 × 640 pixels. The dataset is thoroughly analyzed and all instances of fire and smoke are interpreted. In order to fulfill the experimental prerequisites, the dataset was divided into training, validation, and test sets with an 8:1:1 distribution ratio, resulting in 3180, 397, and 397 images, respectively.

The FLAME dataset was created and released by researchers from Northern Arizona University in partnership with other organizations [41]. The collection comprised 3281 high-resolution aerial images of diverse stages and intensities of forest fires recorded by UAV. In this study, we divided the dataset into training, validation, and test sets in an 8:1:1 ratio to evaluate the performance of our model.

To further verify the effectiveness of the LUFFD-YOLO model, a small UAV remote sensing forest fire (SURSFF) dataset was constructed in this study. The dataset is sourced from two main categories: the first category includes forest fire images captured from UAV viewpoints, collected using web scraping techniques from different platforms; the second category consists of relevant UAV forest fire images obtained from pre-existing forest fire datasets. A total of 110 real-scene UAV images of forest fires were collected with various scales, shapes, and sizes. To mitigate the risk of model overfitting due to the limited dataset size, we employed a series of data augmentation techniques, as illustrated in Figure 1. These techniques included horizontal flipping (Figure 1b), achieved by mirroring the images horizontally using ImageOps.mirror; mosaic data augmentation (Figure 1c), where images were partitioned into four quadrants, each subjected to a random rotation between 0 and 360 degrees and subsequently reassembled to form a mosaic; random directional rotation (Figure 1d), wherein images were rotated by a randomly selected angle within the range of 0 to 360 degrees; noise addition (Figure 1e), in which random noise values ranging from 0 to 50 were applied to each pixel with a random factor, with the resulting pixel values being clipped to the range of 0 to 255. The ultimate dataset was split into 440 training images and 110 validation images, maintaining a ratio of 4:1. The fire category in the expanded dataset was labeled using LabelImg.

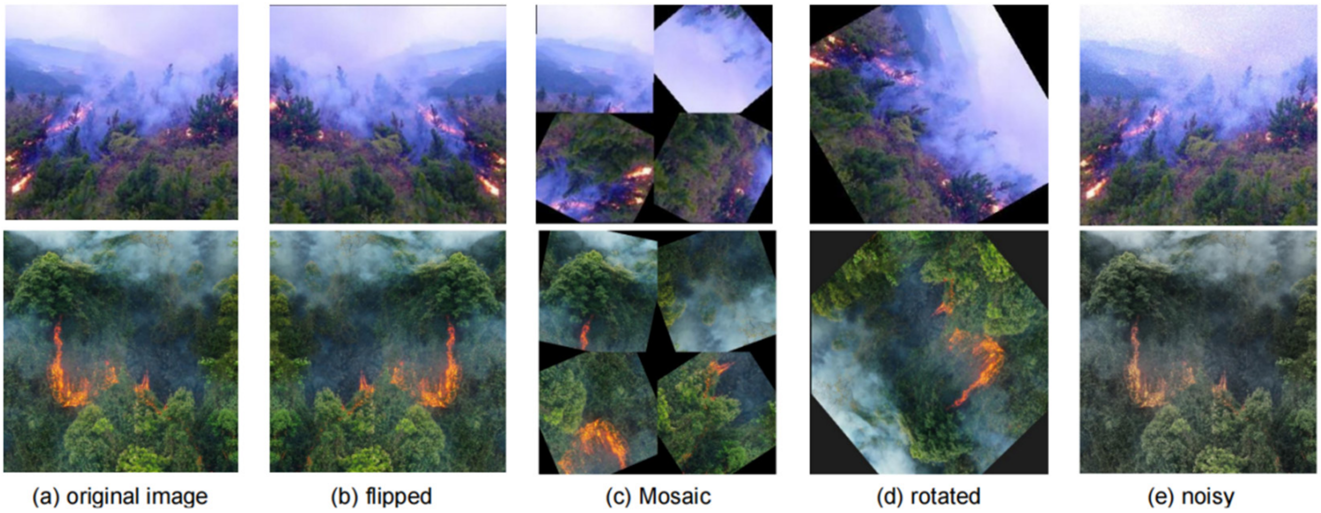

Figure 1. Data augmentation operation: (**a**) original image; (**b**) a horizontal flip operation based on the original image; (**c**) a Mosaic data enhancement operation; (**d**) a random rotation operation; (**e**) random noise addition to the original image.

2.2. Methods

2.2.1. The YOLOv8 Network Architecture

YOLOv8 [42] is a recent addition to the YOLO family, designed specifically for object detection. Depending on the depth and width of its network, it can be categorized into five different frameworks: YOLOv8n, YOLOv8l, YOLOv8s, YOLOv8x, and YOLOv8m. Given the real-time demands of forest fire detection, this study adopts the lightweight YOLOv8n model as the baseline and enhances it. The YOLOv8n architecture consists of four main components: the input layer, backbone layer, neck layer, and output layer, as depicted in Figure 2.

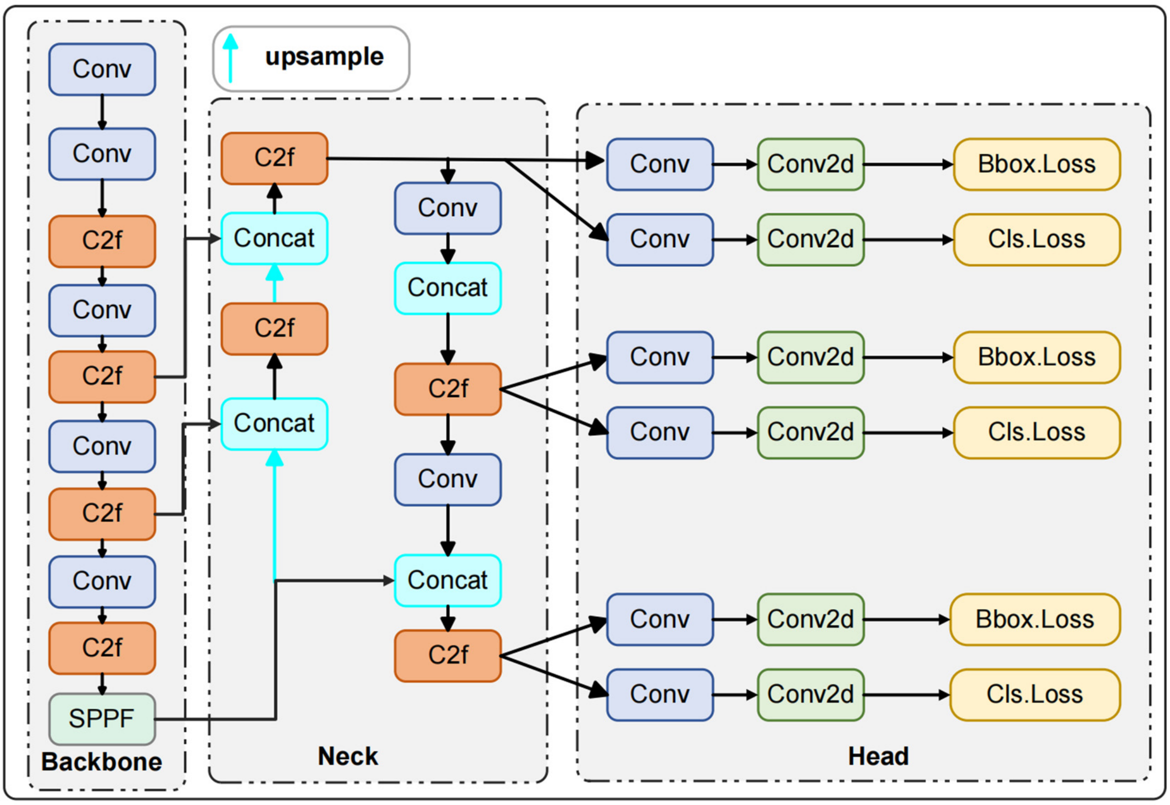

Figure 2. YOLOv8 structure. Note: Conv-convolution; C2f-CSPDarknet53 to 2-Stage feature pyramid networks; Conv2d-convolution with 2-dimension; SPPF-spatial pyramid pooling fusion.

The primary task of the input layer is to receive raw remote-sensing images of forest fires and process them with a series of data augmentation operations to meet the model training requirements. These operations include hue adjustment, image scaling, and the application of Mosaic data augmentation, among others. The principle of Mosaic data augmentation is to randomly select regions from four different images, and then combine these regions into a new image after random cropping and scaling.

The backbone layer is designed to extract essential features from the image, comprising convolution (Conv) layers, CSPDarknet53, 2-Stage feature pyramid networks (C2f), and spatial pyramid pooling fusion (SPPF) modules. The Conv module processes data through convolution operations, batch normalization (BN), and sigmoid linear unit(SiLU) activation functions. The C2f module improves gradient propagation and augments the information flow in the feature extraction network through the integration of cross-layer connections. Unlike the spatial pyramid pooling (SPP) module used in previous YOLO versions, the SPPF module utilizes three consecutive pooling operations to reduce computational complexity while still integrating multi-scale information, thereby expanding the receptive field.

The core function of the neck layer is to achieve cross-dimensional integration of functional features. Through the feature pyramid networks (FPN) and path aggregation network (PAN) structures, it can efficiently merge feature maps of different levels, ensuring the precise preservation of spatial information. This layer enables the model to focus more on target feature information, significantly improving the model's detection performance [43].

The task of the output layer is to produce the final object detection results. It leverages the detailed feature maps produced by the neck layer to determine bounding box positions, category probabilities, and other essential information for each feature map. Furthermore, the output layer employs Non-Maximum Suppression (NMS) [44] technology to remove duplicate detections, maintaining accurate prediction results.

2.2.2. The Proposed LUFFD-YOLO Network

- Lightweight optimization.

Given the importance of promptly detecting fires to prevent their rapid spread, the efficiency of fire detection is a critical factor to consider when evaluating fire detection algorithms. The backbone network of the YOLOv8n model extensively uses convolution operations to increase the number of channels, thereby enlarging the receptive field. However, this increases the model's parameter count and computational cost, which is not conducive to real-time forest fire detection tasks. The main idea behind designing lightweight network structures is to improve the efficiency of detection by adopting more efficient convolutional networks through improving convolution methods. Therefore, this study optimizes the backbone network of YOLOv8n by replacing the standard convolutions in the original network with GhostNetV2 [45] modules, thereby constructing a lightweight feature extractor. While there is a slight decrease in detection accuracy, there is a significant reduction in the model's parameters and computational requirements.

GhostNetV2 combines the decoupled fully connected (DFC) attention mechanism with Ghost modules [46] to maintain performance as much as possible while keeping the structure lightweight. The DFC attention mechanism utilizes dynamic filter capsules to weigh features, highlighting important features and suppressing unimportant information. This mechanism dynamically shifts the model's attention to various features, enhancing its accuracy. From a lightweight perspective, the DFC attention mechanism decomposes the attention map into two fully connected layers and gathers features along the horizontal and vertical directions separately. By separating the horizontal and vertical transformations, the attention module's computational complexity is reduced to $\mathrm{O}\left(\mathrm{H}^2\mathrm{W} + \mathrm{HW}^2\right)$, where W is the width of the feature map and H is the height of the feature map, significantly improving the model's computational efficiency. The specific implementation formulas are shown as Equations (1) and (2).

$$\mathbf{a}'_{hw} = \sum_{h'=1}^{H} F^{H}_{h,h'w} \odot \mathbf{z}_{h'w}, h = 1, 2, \cdots, H, w = 1, 2, \cdots, W, \tag{1}$$

$$\mathbf{a}_{hw} = \sum_{w'=1}^{W} F^{W}_{w,hw'} \odot \mathbf{a}'_{hw'}, h = 1, 2, \cdots, H, w = 1, 2, \cdots, W, \tag{2}$$

where $\mathbf{a}'_{hw}$ and $\mathbf{a}_{hw}$ are the generated attention map of the vertical and horizontal direction $\odot$ is element-wise multiplication, $\boldsymbol{W}$ is the width of the feature map; H is the height of the feature map; F is the learnable weights in the fully connected (FC) layer. $\boldsymbol{F^H}$ and $\boldsymbol{F^W}$ are transformation weights of the vertical and horizontal direction, respectively. Feature $\mathrm{Z} \in \mathbb{R}^{\mathrm{H \times W \times C}}$ can be seen as $\boldsymbol{HW}$ tokens $\mathrm{z_i} \in \mathbb{R}^{\mathrm{C}}$, i.e., $\mathrm{Z} = \{\mathrm{z}_{11}, \mathrm{z}_{12}, \cdots, \mathrm{z}_{\mathrm{HW}}\}$.

GhostNetV2 is primarily divided into configurations with strides of 1 and 2, as shown in Figure 3. In the setup employing a stride of 1, GhostNetV2 incorporates the Inverted bottleneck design from GhostNet alongside the DFC attention mechanism, running concurrently with the initial Ghost module. Subsequently, the Ghost module and DFC attention mechanism are multiplied elementwise to implement the function of enhancing and expanding the features, as illustrated in Figure 4. The enhanced features are subsequently input into the second Ghost module to generate the output features. This process maximizes the capture of long-distance dependencies between pixels at various spatial locations, thereby improving the module's expressive capability.

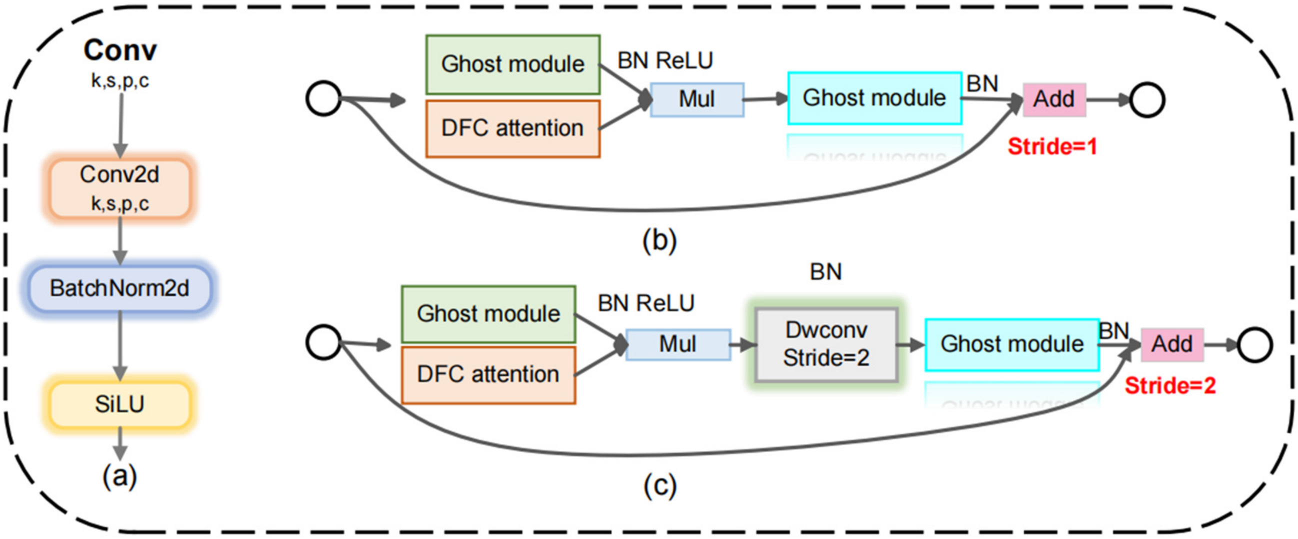

Figure 3. The structure of Conv and GhostNetV2: (**a**) The structure of Conv; (**b**) The structure of GhostNetV2 when the stride is 1; (**c**) The structure of GhostNetV2 when the stride is 2.

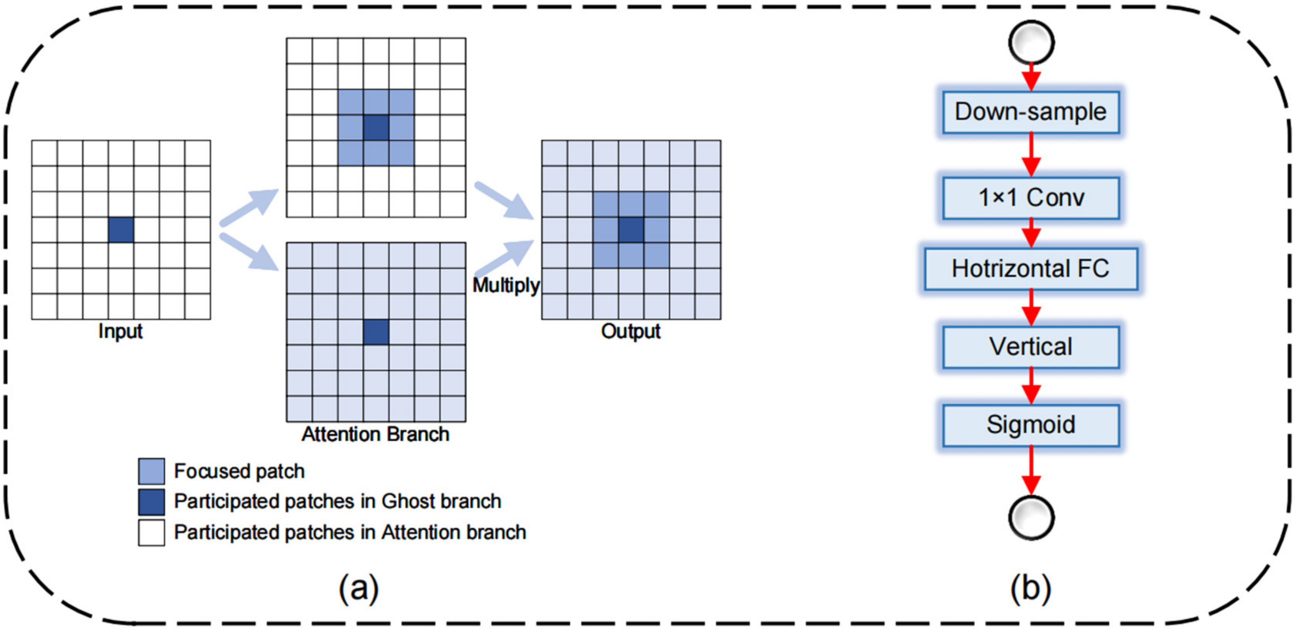

Figure 4. (**a**) Ghost module with DFC attention dot product details; (**b**) DFC attention structure.

In the configuration with a stride of 2 of the GhostNetV2 structure, depth-wise separable convolution is additionally introduced. Following the parallel module of Ghost and DFC attention mechanism, the features are immediately down-sampled to reduce the spatial size of the feature map. This design, while reducing computation and memory usage, also increases the receptive field, facilitating the capture of a broader range of contextual information and minimizing gradient loss. Subsequently, the feature dimension is restored through the Ghost module to ensure consistency with the input.

- Optimization of small forest fire detection using attention mechanisms.

In the context of forest fire monitoring, the presence of small forest fires frequently serves as an indication of fires in their first stages, making it crucial to notice them promptly. Due to its overly simplistic lightweight network structure, the extraction of feature and location information is constrained, hindering the attainment of high accuracy in detecting small forest fires. Thus, this study takes inspiration from attention mechanisms and incorporates the Multi-Head Self Attention (MHSA) mechanism [47] into the C2f structure of the backbone network to introduce the ESDC2f module, as depicted in Figure 5. This module enables the model to learn diverse features in different representational subspaces while fully considering the contextual information of image sub-environments, thereby enhancing the accuracy of small forest fire detection.

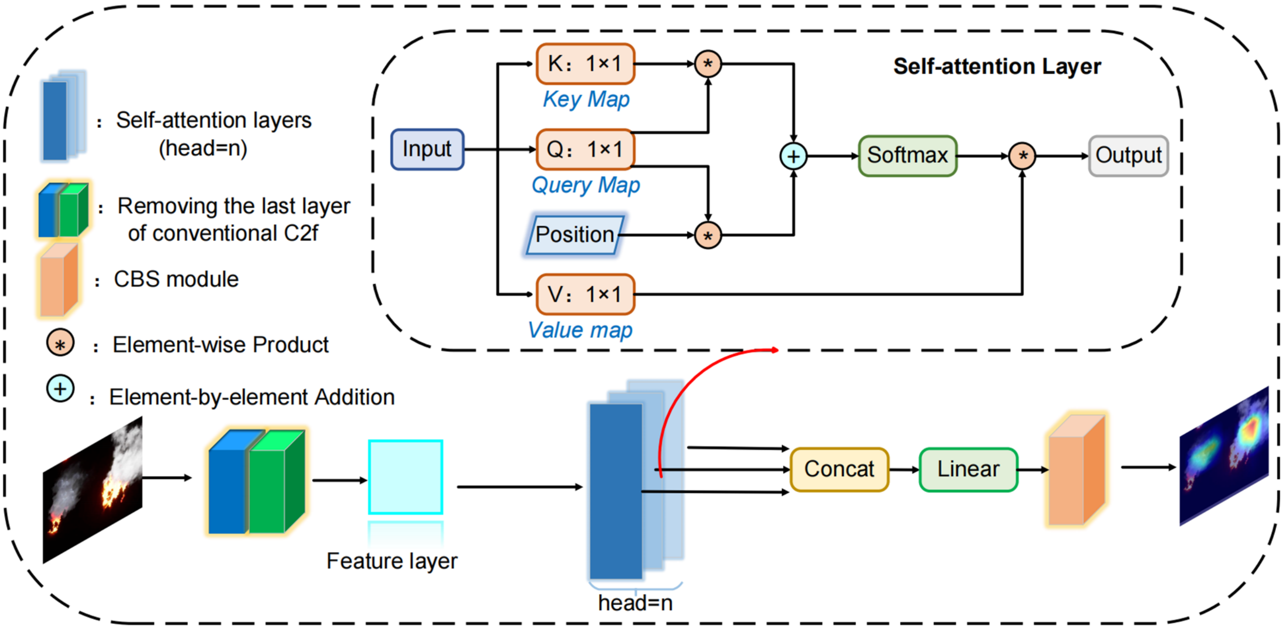

Figure 5. Newly designed ESDC2f structure using the MHSA mechanism. Note: K-Key map; Q-Query map; V-Value map; CBS-Conv2d+BatchNorm2d+SiLU.

Traditional self-attention mechanisms can directly calculate the dependencies between elements within a sequence, effectively capturing the global information of images. However, the information processing capability and perspective of self-attention mechanisms are limited; they learn features only from a single representational subspace. This limits the model's ability to capture different dimensional features of the input data. Therefore, MHSA efficiently captures feature information from different subspaces by using multiple self-attention mechanisms in parallel to process the input data. MHSA first passes the input image features to multiple independent self-attention heads simultaneously. In each head, the model calculates the Query (Q), Key (K), and Value (V) transformations of the input data. Subsequently, each head calculates attention weights based on the similarity between queries and keys. This operation is implemented through a dot product operation followed by a softmax function. The resulting attention weights indicate the importance of each element to other elements within the data. These weights are then used to weight the corresponding values to generate weighted outputs. Finally, the outputs produced by all heads are concatenated and passed through another linear transformation to form the final output. The formulas for Q, K, and V transformations, and the calculation of attention weights are shown as Equations (3)–(6).

$$Q = W_q X, \tag{3}$$

$$K = W_k X, \tag{4}$$

$$V = W_v X, \tag{5}$$

$$Attention(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \cdot V, \tag{6}$$

where Q, K, V are the Query, Key, and Value transformations of the input data X; W_q, W_k, and W_v correspond to the weight matrices of Q, K, and V, respectively; *Softmax* is used to convert the attention scores into a probability distribution. It ensures that the scores are positive and sum up to 1; K^T is the transpose of the Key matrix; d_k (the dimensionality of the keys) is its scaling factor.

- Optimization of forest fire feature extraction capability.

Due to the complexity of forest environments, there can be extreme variations in lighting conditions within forested areas, such as direct sunlight or shadows. Additionally, the diversity of natural landscapes within forests, including variations in tree species, shrubs, and terrain, compound the difficulty of distinguishing forest fires from their natural background. This complexity undeniably imposes more stringent accuracy demands on models designed for forest fire detection. The YOLOv8n model employs the C2f structure in its neck layer to merge semantic information from various layers and scales, enhancing detection performance through efficient feature extraction and cross-layer connections. While this design enhances detection performance, it also increases the model's computational complexity and parameter count. Moreover, its ability to fuse features from different levels remains insufficient. Therefore, this study introduces the SegNeXt Attention [48] mechanism into the traditional C2f of YOLOv8n, proposing the new and efficient plug-and-play HFIC2f module, as shown in Figure 6. This module can automatically emphasize flame features against complex backgrounds and suppress background noise. It also effectively integrates features from different levels, ensuring that while capturing the global shape of forest fires, it can also precisely identify the edges and details of the fires.

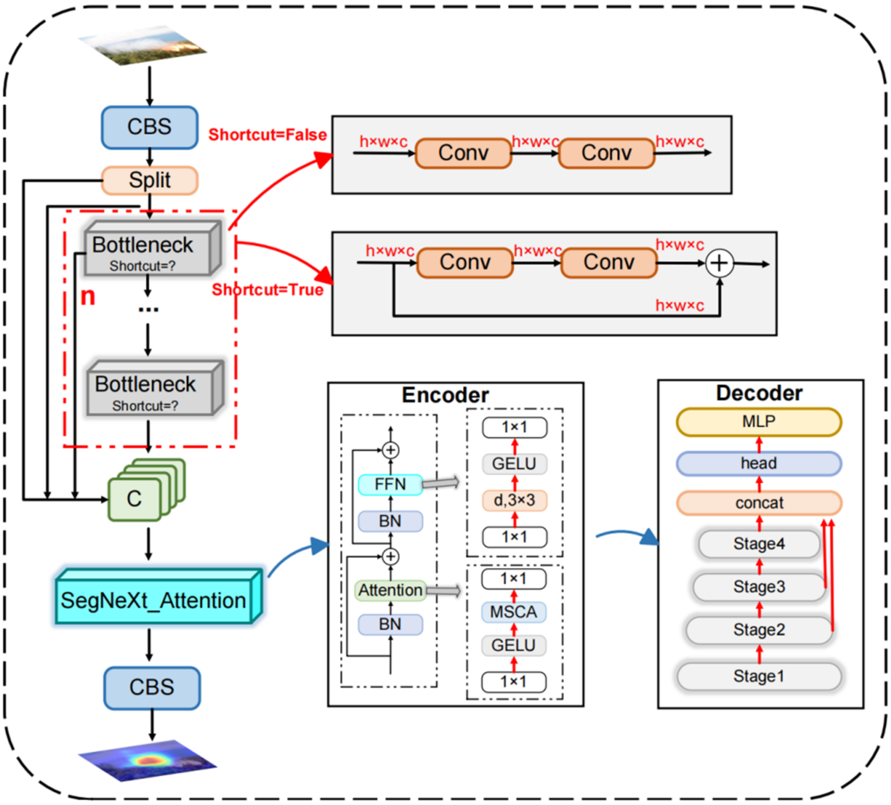

Figure 6. Newly designed HFIC2f structure. Note: MSCA: multi-scale convolutional. attention; MLP: multilayer perceptron.

The SegNeXt attention mechanism improves the model's capacity to process spatial relationships and complex backgrounds in image data by incorporating a multi-scale attention mechanism; its encoder-decoder utilizes a hierarchical progressive approach for image processing, as depicted in Figure 6. During the encoding phase, the input image undergoes gradual down sampling. This is achieved through convolutional and pooling layers, which systematically reduce the image's spatial dimensions. Concurrently, the feature information is enriched and abstracted, enabling the model to glean deeper semantic meaning. Then, the introduced MSCA (Multi-Scale Contextual Attention) module enhances the model's perception of target features by weighting feature maps to highlight important areas and suppress irrelevant backgrounds for more precise target localization and recognition. MSCA comprises three components: depth-wise convolution for aggregating local information, multi-branch depth-wise dilated convolution for capturing multi-scale context, and 1 × 1 convolution for modeling inter-channel relationships. The specific mathematical representation is shown as Equations (7) and (8).

$$Att = Conv_{1\times1}\left(\sum_{i=0}^{3} Scale_i(DW\text{-}Conv(F))\right), \tag{7}$$

$$Out = Att \ \otimes F.\,, \tag{8}$$

where *Att* and *Out* are the attention map and output, respectively; *F* represents the input feature; ⊗ is the element-wise matrix multiplication operation; DW-Conv denotes depth-wise convolution; $Scale_{\mathrm{i}}$ is the identity connection, $\mathrm{i} \in \{0,1,2,3\}$, denotes the *i*th branch.

In the decoding phase, the model gradually restores the spatial resolution of the image through a series of up sampling operations and convolutional layers, utilizing deep features obtained during the encoding phase. Feature fusion techniques are used at this stage. They integrate deep semantic information from the encoding phase with high-resolution features from the decoding phase, enhancing the detail of reconstructed feature maps. During this process, the unique cross-layer connections of SegNeXt Attention ensure that even in cases of rich image details or complex backgrounds, the model can effectively determine the target areas. Finally, through the high-resolution feature maps output by the decoder, the model generates accurate results, achieving efficient detection and recognition of forest fires in complex backgrounds. Simultaneously, to ensure the decoder's lightness, features from the last three stages (stage2, stage3 and stage4 in Figure 6) are aggregated, using a lightweight "hamburger" to further model the global context, maintaining high computational efficiency while still being lightweight.

2.3. Experimental Setup and Accuracy Assessment

The experiment is performed on a 64-bit Windows 10 system, using Python and the PyTorch 2.3.0 library for evaluation. The setup also includes an NVIDIA GeForce RTX 4090 graphics card with 24 GB of VRAM, offering substantial graphics processing power. For consistent experimental procedures and comparable outcomes, this configuration will be kept consistent throughout subsequent parts of the study.

To comprehensively and effectively evaluate forest fire detection models, four metrics were employed: precision (*P*), recall (*R*), *F*1 score, and mean Average Precision (*mAP*). Below are the respective calculation formulas (Equations (9)–(12)).

$$P = \frac{TP}{TP + FP}, \tag{9}$$

$$R = \frac{TP}{TP + FN}, \tag{10}$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R}, \tag{11}$$

$$mAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q}, \tag{12}$$

where *TP* denotes the number of true positives, representing the correctly detected forest fires by the models. *FP* represents false positives, indicating instances where models detect non-existent fires. *FN* stands for false negatives, referring to forest fires missed by the models. Precision (P) quantifies the proportion of actual forest fires accurately detected, while recall (R) measures the proportion of correctly identified forest fires out of all actual positive samples. The $F1$ score, the harmonic mean of precision and recall, offers a balanced assessment of the model's performance. Mean Average Precision (mAP) is calculated by averaging the Average Precision (AP) across different categories, serving as a comprehensive metric to evaluate a detection model's performance over the entire dataset.

In addition, this study calculates the number of model parameters (including weights and biases) and floating-point operations per second (FLOPs) to further measure the complexity and computational requirements of a model. The number of FLOPs directly determines the training and inference speed of the deep learning model. The more FLOPs, the higher the computational demand of the model and the more stringent the hardware requirements.

3. Results

3.1. Comparison between YOLOv8n and LUFFD-YOLO

To assess the efficacy of the proposed LUFFD-YOLO model, we performed a comparative study against the YOLOv8n model using the extensive public forest fire dataset (M4SFWD). Figure 7a–c represents the discrepancies between the predicted bounding boxes, classifications, and probability distributions compared to the true values. The closer the curves are to zero, the smaller the difference from the true values. From Figure 7a–c, it can be observed that the losses for LUFFD-YOLO are closer to zero, which indicates a better training effect. The *mAP50* value refers to the average precision of a model in correctly identifying positive objects when the Intersection over the Union (IoU) threshold is set at 50%. The *mAP50-95* value is the average of the mean Average Precision values calculated at different IoU thresholds, typically from 50% to 95% in increments of 5%. From the curves shown in Figure 7d–g, it is apparent that LUFFD-YOLO maintains higher values compared to YOLOV8n across four different metrics—precision, recall, *mAP50*, and *mAP50-95*—as the epochs progress.

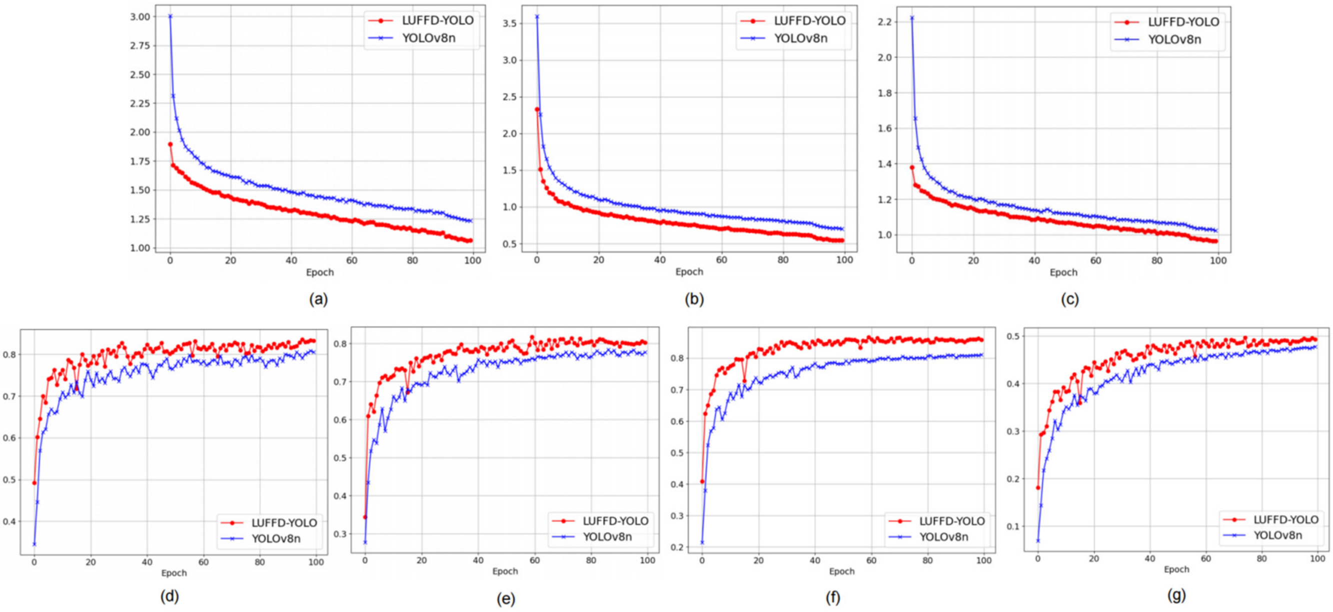

Figure 7. Comparison of training results of LUFFD-YOLO and YOLOv8n. (**a**) box loss; (**b**) classification loss; (**c**) distribution focal loss; (**d**) precision curve; (**e**) recall curve; (**f**) *mAP50* curve; (**g**) *mAP50-95* curve.

The experimental findings, depicted in Table 1, demonstrate that the LUFFD-YOLO model surpasses the baseline model across all four evaluation metrics (precision, recall, $F1$ score, and mAP), exhibiting enhancements of 4.2%, 7.2%, 5.7%, and 5.1%, respectively.

This is primarily attributed to the newly designed backbone layer structure, ESDC2f, which enhances the model's capability to detect small-scale forest fires. The addition of the HFIC2f structure in the neck layer improves the model's ability to detect forest fires in the presence of complicated backdrops accurately. The decrease in the number of parameters is mainly attributed to substituting conventional convolutional layers with GhostNetV2. In summary, the proposed LUFFD-YOLO model significantly enhances the accuracy of forest fire detection while maintaining a lightweight design, which is of great significance for the timely monitoring of forest fires.

Table 1. Comparison of the YOLOv8n and LUFFD-YOLO on the test data of the M4SFWD dataset.

Model	Precision (%)	Recall (%)	*F*1 (%)	*mAP* (%)	Parameters (M)
YOLOv8n	77.6	75.6	76.6	82.1	3.0
LUFFD-YOLO	80.9	81.1	81.0	86.3	2.6

Figure 8 illustrates the forest fire detection results of YOLOv8n and LUFFD-YOLO across different scenarios: (from left to right) low-light forest fire, dense forest small fire, low-light snowfield, multi-target forest fire on an island, and high-brightness forest fire. The detection accuracy of LUFFD-YOLO across various scenarios (Figure 8(c-1–5)) is superior to that of YOLOv8n (Figure 8(b-1–5)).

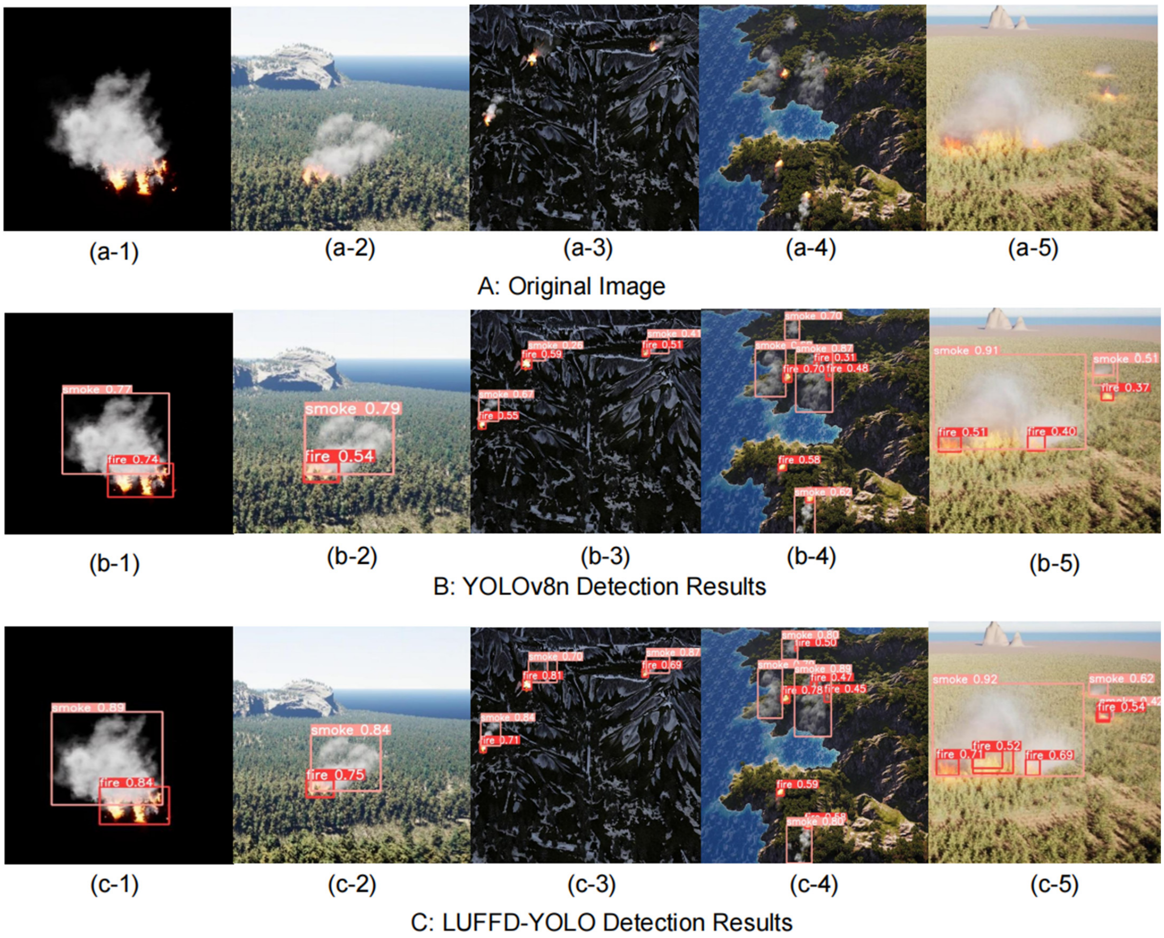

Figure 8. Visual comparison of YOLOv8n and LUFFD-YOLO detection accuracy for different scenes. The "a" to "c" represents an original image, the detection result using the YOLOv8n model and the detection result using LUFFD-YOLO, respectively; the "1" to "5" represent low-light forest fire, dense forest small fire, low-light snowfield, multi-target forest fire on an island, and high-brightness forest fire scenes, respectively.

3.2. Ablation Experiment

To further assess the detection performance of the proposed LUFFD-YOLO model, we conducted ablation studies to evaluate the individual impact of each enhancement step (model lightweight, multi-head attention mechanism, SegNeXt Attention) on the model's performance. The accuracy metrics of the ablation experiment include precision, recall, *F1*, *mAP*, and FLOPs.

Table 2 displays the ablation experimental results using the M4SFWD dataset under different optimization measures. The results from Methods (1) indicated that when optimizing the backbone network with GhostNetV2, the model's FLOPs were reduced by 14.8%. This reduction is mainly due to the introduction of a more efficient feature generation mechanism by GhostNetV2, which generates original features using a minimal number of basic convolution operations and then expands them into more features through cost-effective operations, thereby reducing computational costs. Additionally, it further enhances the model's computational efficiency and performance through improved feature reuse and feature fusion strategies.

Table 2. Results of ablation experiments on the M4SFWD dataset.

Name	Models	Precision (%)	Recall (%)	*F1* (%)	*mAP* (%)	FLOPs (G)
YOLOv8n	YOLOv8n (baseline)	77.6	75.6	76.6	82.1	8.1
Methods (1)	YOLOv8n+GN	75.8	74.2	75.0	80.3	6.9
Methods (2)	YOLOv8n+GN+ESDC2f	78.8	79.1	78.9	84.5	7.0
Methods (3) (ours)	YOLOv8n+GN+ESDC2+HFIC2f	80.9	81.1	81.0	86.3	7.0

Note: GN-GhostNetV2; ESDC2f-enhanced small-object forest fire detection C2f; HFIC2f-hierarchical feature-integrated C2f.

However, the lightweight structure can also lead to a slight decrease in detection accuracy. To compensate for the accuracy loss caused by lightweighting, Method (2) incorporates the ESDC2f structure into the backbone network, resulting in respective improvements in precision, recall, and *mAP* of 4.0%, 6.6%, and 5.2%, compared to Method (1). Figure 9 presents the first eight feature maps of the same layer generated by Methods (1) and Methods (2). It can be observed the feature maps generated by Methods (2), which integrate the GhostNetV2 and ESDC2f structure into the backbone network, demonstrate superior globality and extensibility compared to Methods (1), which solely incorporate the GhostNetV2 structure. This suggests that the feature maps of Methods (2) exhibit greater diversity and contain a more abundant amount of information. The primary enhancement stems from the MHSA attention mechanism, enabling the model to capture varied features across different representational subspaces while comprehensively considering contextual information from image sub-environments. This contributes significantly to improving the accuracy of forest fire detection. At the same time, it is observed that the model still maintains a reduced computational cost after the addition of this structure. Methods (3) demonstrate that upon adding the newly designed HFIC2f structure to the Neck layer on top of Methods (2), the model's precision, recall, and *mAP* metrics significantly increased compared to the baseline model YOLOv8n, with increases of 4.3%, 7.3%, and 5.1%, respectively. This is attributed to the module's ability to better highlight flame features in complex environments and reduce background noise. It effectively integrates features at different levels, capturing both the overall shape of forest fires and precisely identifying the edges and details of forest fires. LUFFD-YOLO ultimately adopts the structure from Methods (3) to ensure optimal model performance, achieving a *mAP* of 86.3% on the M4SFWD dataset with a computational cost of 7.0 G. Compared to the baseline model YOLOv8n, it maintains high detection accuracy while being lightweight.

Additionally, we employed Grad-CAM for interpretative analysis of the model's improvement strategies. Figure 10 displays the heatmaps of forest fire detection on the M4SFWD dataset using both the baseline YOLOv8n model and the proposed LUFFD-YOLO model. The figure demonstrates that LUFFD-YOLO detects larger target areas for both high-brightness and low-light forest fires compared to YOLOv8n, as seen in comparisons between Figure 10(b-1,2) with Figure 10(c-1,2). Moreover, in situations with complex environments and irregularly sized targets, LUFFD-YOLO better focuses on the positive sample areas of forest fires and smoke, effectively overcoming the problem of missed detections (compared Figure 10(b-3,4) with Figure 10(c-3,4)). This indicates that the introduction of the ESDC2f structure enables the model to more fully capture image features, enhancing feature representation. Simultaneously, the introduction of the HFIC2f structure significantly improves the model's perception of targets against complex backgrounds. The experimental results prove the effectiveness of the improvement measures.

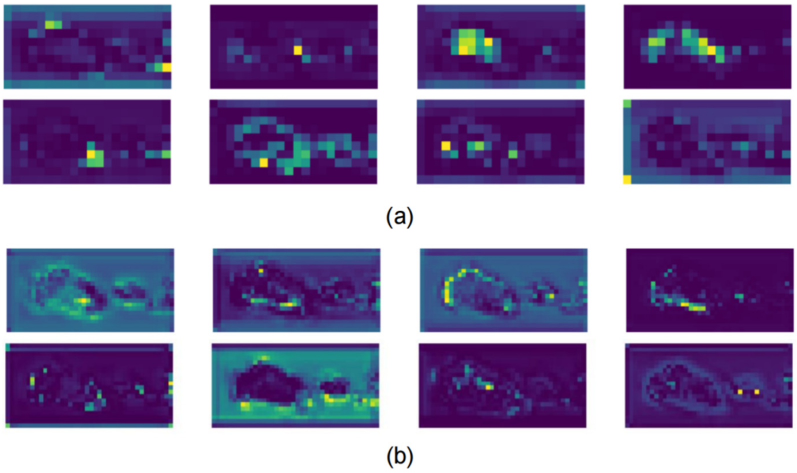

Figure 9. Feature map visualization: (**a**) Methods (1): baseline+GN; (**b**) Methods (2): baseline+GN+ESDC2f.

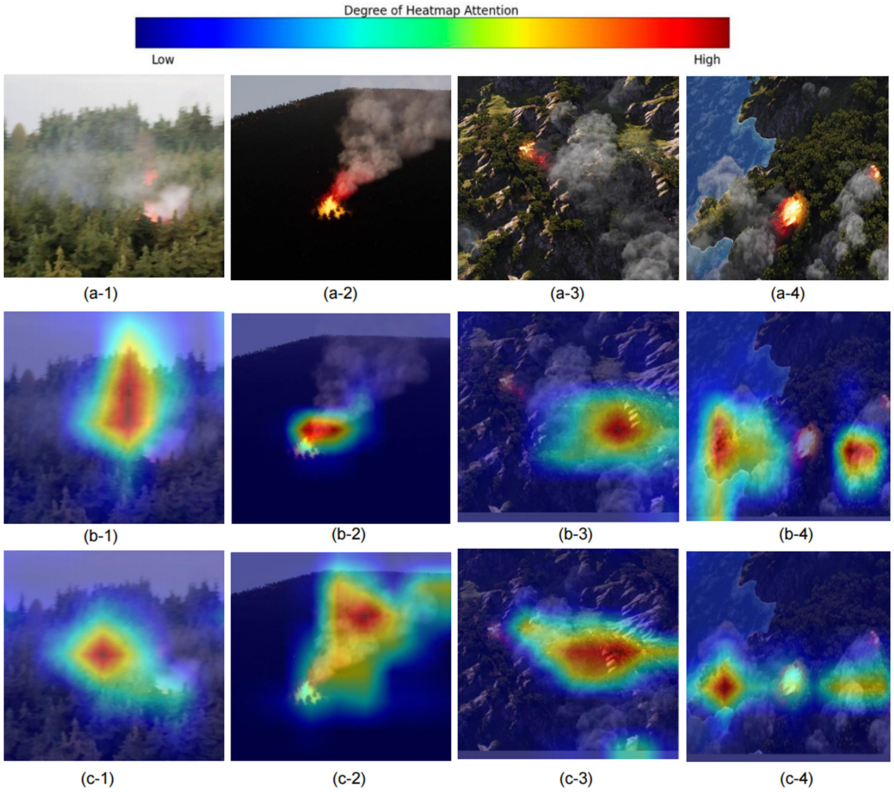

Figure 10. Grad-CAM visualization: The "a" to "c" represents an original image, YOLOv8n model, and LUFFD-YOLO model, respectively; the "1" to "4" represent high-brightness forest fire, low-light forest fire, a small fire in a mountain area, and multi-target forest fire on an island, respectively.

3.3. Verification Experiment

To verify the performance of the LUFFD-YOLO model, we compared the model with four lightweight models: YOLOv3-tiny [49], YOLOv5 [50], YOLOv7-tiny [51], and YOLOv8n using the new data (FLAME and SURSFF dataset). Table 3 shows that the LUFFD-YOLO model achieved the best accuracy with the minimum parameters on both the FLAME dataset (precision: 87.1%, recall: 87.5%, *mAP*: 90.1%, *M*: 2.6) and the SURSFF dataset (precision: 88.9%, recall: 86.7%, *mAP*: 90.9%, *M*: 2.6). It shows that the *mAP* of LUFFD-YOLO increased by 5.5%, 5.3%, 28.7%, and 3.1% on the FLAME dataset, and by 5.2%, 2.2%, 4.3%, and 3.5% on the SURSFF dataset when compared to YOLOv3-tiny, YOLOv5, YOLOv7-tiny, and YOLOv8n, respectively. For the FLAME dataset, the precision of the LUFFD-YOLO model increased by 3.8%, 4.7%, 16.9%, and 2.7%, and the recall increased by 12.1%, 8.8%, 22.9%, and 5.8% comparing to YOLOv3-tiny, YOLOv5, YOLOv7-tiny, and YOLOv8n, respectively. Similarly, for the SURSFF dataset, the precision of LUFFD-YOLO increased by 6.3%, 2.1%, 4.5%, and 3.7%, and the recall increased by 6.1%, 2.0%, 4.5%, and 3.0% comparing to YOLOv3-tiny, YOLOv5, YOLOv7-tiny, and YOLOv8n, respectively. It indicated that the LUFFD-YOLO model dramatically reduced both commission and omission errors when detecting forest fires. In addition, LUFFD-YOLO has the smallest number of parameters, which makes it highly beneficial for real-time forest fire detection tasks.

Table 3. Comparison of different models on the FLAME and SURSFF Datasets.

Dataset	Model	Precision (%)	Recall (%)	*mAP* (%)	Parameters (M)
FLAME	YOLOv3-tiny	83.9	78.0	85.4	8.1
	YOLOv5	83.2	80.4	85.6	47.1
	YOLOv7-tiny	74.5	72.8	70.0	6.0
	YOLOv8n	84.8	82.7	87.4	3.0
	LUFFD-YOLO	87.1	87.5	90.1	2.6
SURSFF	YOLOv3-tiny	83.9	81.4	86.4	8.1
	YOLOv5	87.2	84.9	88.9	47.1
	YOLOv7-tiny	85.2	82.9	87.1	6.0
	YOLOv8n	86.4	83.5	87.8	3.0
	LUFFD-YOLO	88.9	86.7	90.9	2.6

Figure 11 visualizes the results of various models on the FLAME and the SURSFF dataset. The numerical values of the predicted bounding boxes in the figure represent the confidence scores of forest fires. Confidence is a value between 0 and 1 that indicates the model's certainty in the detected target's presence. A confidence close to 1 suggests the model is confident that the target object is contained within the box, whereas a confidence close to 0 suggests the model believes the box likely does not contain the target. It can be observed that LUFFD-YOLO maintains higher detection accuracy. This further indicates that the model proposed in this article can better sustain detection performance through network lightweight.

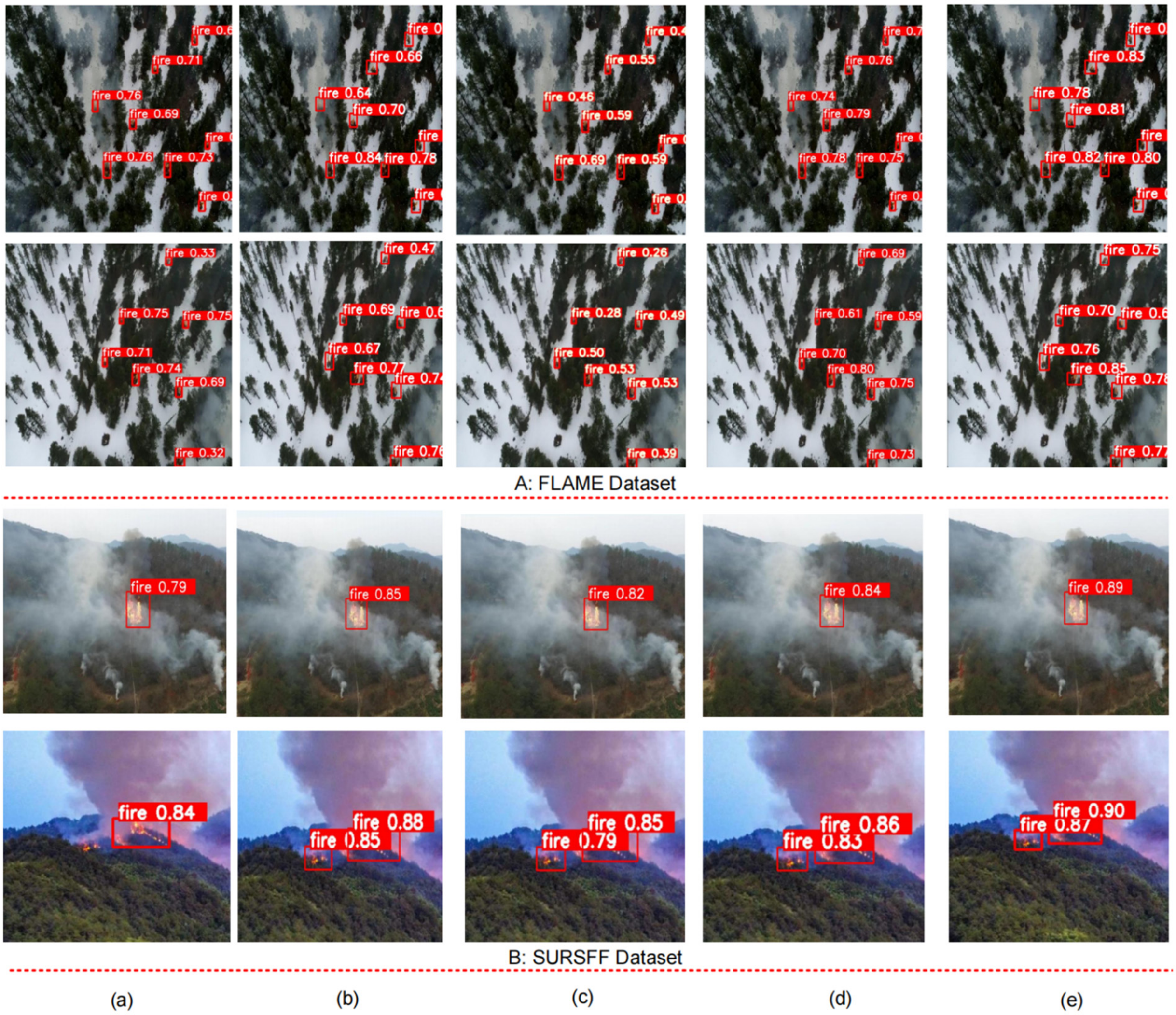

Figure 11. Visual comparison of detection results between LUFFD-YOLO and four lightweight models using the FLAME and SURSFF dataset: (**a**) YOLOv3-tiny; (**b**) YOLOv5; (**c**) YOLOv7-tiny; (**d**) YOLOv8n; (**e**) LUFFD-YOLO. Note: The numerical values of the predicted bounding boxes in the figure represent the confidence scores of forest fires. Confidence is a value between 0 and 1 that indicates the model's certainty in the detected target's presence.

4. Discussion

4.1. The Advantages of the Proposed LUFFD-YOLO Model

This study proposes LUFFD-YOLO, a lightweight object detection model that achieves a high level of accuracy while maintaining a good balance between its architecture and detection performance. The baseline YOLOv8 model demonstrates exceptional performance in the field of one-stage object detection. However, to attain optimal detection performance, the YOLOv8 model employs a significant amount of convolutional operations, which include numerous inefficient computations. This undoubtedly increases the model's computational load, which hinders real-time detection tasks for forest fire targets. Therefore, this study utilizes the GhostNetV2 architecture to enhance the conventional convolutions in the backbone layer of the YOLOv8n model. This optimization enhances computational efficiency and decreases the number of parameters, while still maintaining good performance. Additionally, this study proposes a new structure, ESDC2f, with the aid of the MHSA attention mechanism. This structure enables the model to learn diverse features in different representational subspaces and fully consider the contextual information of the image sub-environment, thereby enhancing the detection accuracy of different target scales. The ESDC2f structure significantly enhances the feature extraction capabilities of the backbone layer, in comparison to the YOLOv8n model. Finally, drawing inspiration from SegNeXt's attention, a new structure, HFIC2f is proposed. It emphasizes target features against complex backgrounds, suppresses noise, and enhances the model's ability to recognize forest fire-related features in complex environments. The addition of the HFIC2f structure in LUFFD-YOLO has significantly improved the ability to integrate features at various levels, as compared to the conventional neck layer of YOLOv8n.

4.2. Comparative Experiments of Different Models

Based on the M4SFWD dataset, the proposed one-stage LUFFD-YOLO model was compared with traditional two-stage object detection models (Faster-RCNN [52] and SSD [53]) and lightweight one-stage object detection algorithms (YOLOv3-tiny [49], YOLOv5 [50], YOLOv6s [54], YOLOv7-tiny [51]). Table 4 indicates that LUFFD-YOLO surpasses the comparison algorithms across all performance metrics. Specifically, Faster-RCNN, which extracts features from the entire image before region proposal, suffers from low precision in forest fire detection owing to its intricate model structure, failing to satisfy real-time detection needs (lowest *F*1 and a large number of parameters). Although the SSD model exhibits higher detection accuracy (*F*1: 70.6%), its reliance on fixed-size anchor boxes restricts its capability to capture the diversity and variability of forest fires. It also has the largest number of parameters (560.6 M), indicating the lowest efficiency. The YOLOv3-tiny model, by simplifying the network structure to reduce computational demands (parameters: only 8.1 M), compromises its ability to capture complex features of forest fires but provides a moderate accuracy (*F*1: 72.0%). In contrast, YOLOv5, with its auto-learned anchor sizes and PANet feature fusion technology, improves detection capabilities (*F*1: 72.9%; *mAP*: 79.3%) but its deep and complex network structure is unsuitable for real-time detection tasks (parameters: 47.1M). The YOLOv6s model enhances feature fusion efficiency and model convergence speed through the Rep-PAN structure and the decoupling strategy of YOLOX's detection head, providing an *F*1 of 73.5% with parameters of 17.2 M. The YOLOv7-tiny model, adopting the efficient layer attention network structure, effectively aggregates features at different levels, enhancing the network's feature extraction capability, while optimizing the computational process with the MaxPool2d structure to speed up detection, yet the model's weight size remains 6.0 M. The YOLOv7-tiny achieves the highest accuracy and lowest parameters among the compared models (*F*1: 74.4%; *mAP*: 80.4%; parameters: 6.0 M). In comparison, LUFFD-YOLO not only optimizes detection performance (*F*1: 81%; *mAP*: 88.3%) and achieves model lightweight (parameters: 2.6 M) but also enhances the ability to recognize targets of varying sizes, offering a new effective methodology for forest fire detection.

Table 4. Comparative experiments with different models on the M4SFWD dataset.

Model	Precision (%)	Recall (%)	*F*1 (%)	*mAP* (%)	Parameters (M)
Faster-RCNN	69.4	68.3	68.9	76.8	120.4
SSD	70.6	70.5	70.6	77.5	560.6
YOLOv3-tiny	72.7	71.4	72.0	78.1	8.1
YOLOv5	73.1	72.7	72.9	79.3	47.1
YOLOv6s	74.2	72.9	73.5	79.9	17.2
YOLOv7-tiny	75.4	73.5	74.4	80.4	6.0
LUFFD-YOLO	80.9	81.1	81.0	88.3	2.6

4.3. Limitations and Future Work

While the LUFFD-YOLO model showcased notable advantages in forest fire detection within this study, it still carries certain limitations. Firstly, the GHostNetV2 structure employed in this study uses cost-effective operations to reduce the model's parameter count and improve computational efficiency. However, these cost-effective operations have a limited receptive field and can only capture local, fine-grained features, which limits the model's feature extraction capability. In future research, the model architecture could be further enhanced to improve the capability of extracting features from lightweight structures. For instance, it might be beneficial to disregard the connections between structures in order to produce a more condensed model structure. Additionally, convolutional kernels of various sizes can be utilized to effectively capture multi-scale features.

Secondly, small forest fires only occupy a limited number of pixels in the entire image, resulting in a loss of clarity and definition in intricate characteristics such as shape, color, texture, and so on. Additionally, small forest fires have a sparse feature representation and a low signal-to-noise ratio, which means that their feature signals are readily overshadowed by background noise. Super-resolution technology-based methods provide robust interpretability for small object detection. Thus, by utilizing super-resolution technology to enhance the features of small targets and adopting higher quality features, we aim to further improve the model's perception ability for small target detection.

Finally, this study only focused on the UAV-borne RGB images, which means that crucial information like the temperature of objects can not be captured. Thus, with the availability of UAV-borne thermal infrared sensors, a deep learning algorithm that integrates RGB images with thermal infrared images should be developed in the future to improve the detection of forest fires by incorporating temperature.

5. Conclusions

This study proposes a lightweight forest fire detection model, LUFFD-YOLO, for UAV remote sensing data. The LUFFD-YOLO model improves the existing YOLOv8n network by incorporating a set of optimizations: Initially, GhostNetv2 was utilized to enhance the conventional convolution of YOLOv8n, resulting in a significant reduction in the model's parameter count. Second, the ESDC2f architecture, which utilizes the MHSA attention mechanism, was proposed. This design enhances the backbone layer's feature extraction capability, thereby improving the accuracy of small forest fire detection. Third, the HFIC2f structure was redesigned by using the SegNeXt attention mechanism to enhance the integration of features from different layers, resulting in improved accuracy in detecting wildfires in complicated backgrounds. The LUFFD-YOLO model shows improvements in precision, recall, and *mAP* of 4.3%, 7.3%, and 5.1%, respectively, compared to the baseline model YOLOv8n, with a reduction in parameter count of 13.3% and desirable generalization on different datasets. This indicates that the model is able to balance between a high level of accuracy in detecting forest fires and model efficiency, ensuring that it can work in real time. However, while reducing the number of parameters in the backbone layer, the model loses a considerable amount of feature information, and its performance in detecting small target forest fires still has shortcomings. In the future, we will keep refining the model, focusing on improving the feature extraction capability of the lightweight structure and the ability to detect small target forest fires to achieve better forest fire detection performance. This work will greatly aid real-time forest fire detection using UAV remote sensing images.

Author Contributions: Conceptualization, Y.H. and Z.Z.; methodology, Y.H.; software, Y.H.; validation, Y.H., Z.Z. and B.D.; formal analysis, Y.H. and B.D.; investigation, Y.H. and B.D.; data curation, Y.H., B.D. and R.G.; writing—original draft preparation, Y.H.; writing—review and editing, Y.H., G.Y. and Z.Z.; visualization, Y.H., R.G. and B.D.; supervision, Y.H. and Z.Z.; funding, G.Y. and Y.H.; project administration, G.Y. and Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China under Grant No. 2023YFC3006900 and the National College Students Innovation and Entrepreneurship Training Program (DC-2024471).

Data Availability Statement: The data pertinent to this research are available from the corresponding authors upon request. These data are not publicly accessible as they are derived from lab results.

Acknowledgments: The authors appreciate the editors and anonymous reviewers for their valuable recommendations.

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

References

1. Flannigan, M.D.; Stocks, B.J.; Wotton, B.M. Climate Change and Forest Fires. *Sci. Total Environ.* **2000**, *262*, 221–229. [CrossRef]
2. Flannigan, M.D.; Amiro, B.D.; Logan, K.A.; Stocks, B.J.; Wotton, B.M. Forest Fires and Climate Change in the 21ST Century. *Mitig. Adapt. Strat. Glob. Chang.* **2006**, *11*, 847–859. [CrossRef]
3. Stocks, B.J.; Mason, J.A.; Todd, J.B.; Bosch, E.M.; Wotton, B.M.; Amiro, B.D.; Flannigan, M.D.; Hirsch, K.G.; Logan, K.A.; Martell, D.L.; et al. Large Forest Fires in Canada, 1959–1997. *J. Geophys. Res.* **2002**, *107*, FFR 5-1–FFR 5-12. [CrossRef]
4. Crist, M.R. Rethinking the Focus on Forest Fires in Federal Wildland Fire Management: Landscape Patterns and Trends of Non-Forest and Forest Burned Area. *J. Environ. Manag.* **2023**, *327*, 116718. [CrossRef] [PubMed]
5. Grünig, M.; Seidl, R.; Senf, C. Increasing Aridity Causes Larger and More Severe Forest Fires across Europe. *Glob. Chang. Biol.* **2023**, *29*, 1648–1659. [CrossRef] [PubMed]
6. Hillayová, M.K.; Holécy, J.; Korísteková, K.; Bakšová, M.; Ostrihoň, M.; Škvarenina, J. Ongoing Climatic Change Increases the Risk of Wildfires. Case Study: Carpathian Spruce Forests. *J. Environ. Manag.* **2023**, *337*, 117620. [CrossRef]
7. Turco, M.; Abatzoglou, J.T.; Herrera, S.; Zhuang, Y.; Jerez, S.; Lucas, D.D.; AghaKouchak, A.; Cvijanovic, I. Anthropogenic Climate Change Impacts Exacerbate Summer Forest Fires in California. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2213815120. [CrossRef]
8. Howell, A.N.; Belmont, E.L.; McAllister, S.S.; Finney, M.A. An Investigation of Oxygen Availability in Spreading Fires. *Fire Technol.* **2023**, *59*, 2147–2176. [CrossRef]
9. Menut, L.; Cholakian, A.; Siour, G.; Lapere, R.; Pennel, R.; Mailler, S.; Bessagnet, B. Impact of Landes Forest Fires on Air Quality in France during the 2022 Summer. *Atmos. Chem. Phys.* **2023**, *23*, 7281–7296. [CrossRef]
10. Chen, Y.-J.; Lai, Y.-S.; Lin, Y.-H. BIM-Based Augmented Reality Inspection and Maintenance of Fire Safety Equipment. *Autom. Constr.* **2020**, *110*, 103041. [CrossRef]
11. Sharma, A.; Singh, P.K.; Kumar, Y. An Integrated Fire Detection System Using IoT and Image Processing Technique for Smart Cities. *Sustain. Cities Soc.* **2020**, *61*, 102332. [CrossRef]
12. Wooster, M.J.; Roberts, G.J.; Giglio, L.; Roy, D.P.; Freeborn, P.H.; Boschetti, L.; Justice, C.; Ichoku, C.; Schroeder, W.; Davies, D. Satellite Remote Sensing of Active Fires: History and Current Status, Applications and Future Requirements. *Remote Sens. Environ.* **2021**, *267*, 112694. [CrossRef]
13. Hua, L.; Shao, G. The Progress of Operational Forest Fire Monitoring with Infrared Remote Sensing. *J. For. Res.* **2017**, *28*, 215–229. [CrossRef]
14. Bessho, K.; Date, K.; Hayashi, M.; Ikeda, A.; Imai, T.; Inoue, H.; Kumagai, Y.; Miyakawa, T.; Murata, H.; Ohno, T. An Introduction to Himawari-8/9—Japan's New-Generation Geostationary Meteorological Satellites. *J. Meteorol. Soc. Jpn. Ser. II* **2016**, *94*, 151–183. [CrossRef]
15. Justice, C.O.; Giglio, L.; Korontzi, S.; Owens, J.; Morisette, J.T.; Roy, D.; Descloitres, J.; Alleaume, S.; Petitcolin, F.; Kaufman, Y. The MODIS Fire Products. *Remote Sens. Environ.* **2002**, *83*, 244–262. [CrossRef]
16. Osco, L.P.; Junior, J.M.; Ramos, A.P.M.; de Castro Jorge, L.A.; Fatholahi, S.N.; de Andrade Silva, J.; Matsubara, E.T.; Pistori, H.; Gonçalves, W.N.; Li, J. A Review on Deep Learning in UAV Remote Sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102456. [CrossRef]
17. Li, C.; Li, G.; Song, Y.; He, Q.; Tian, Z.; Xu, H.; Liu, X. Fast Forest Fire Detection and Segmentation Application for UAV-Assisted Mobile Edge Computing System. *IEEE Internet Things J.* **2023**. [CrossRef]
18. Yang, X.; Hua, Z.; Zhang, L.; Fan, X.; Zhang, F.; Ye, Q.; Fu, L. Preferred Vector Machine for Forest Fire Detection. *Pattern Recognit.* **2023**, *143*, 109722. [CrossRef]
19. Maeda, N.; Tonooka, H. Early Stage Forest Fire Detection from Himawari-8 AHI Images Using a Modified MOD14 Algorithm Combined with Machine Learning. *Sensors* **2022**, *23*, 210. [CrossRef]
20. Liu, J.; Guan, R.; Li, Z.; Zhang, J.; Hu, Y.; Wang, X. Adaptive Multi-Feature Fusion Graph Convolutional Network for Hyperspectral Image Classification. *Remote Sens.* **2023**, *15*, 5483. [CrossRef]
21. Guan, R.; Li, Z.; Li, X.; Tang, C. Pixel-Superpixel Contrastive Learning and Pseudo-Label Correction for Hyperspectral Image Clustering. In Proceedings of the ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 6795–6799.
22. 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]
23. Liu, F.; Chen, R.; Zhang, J.; Xing, K.; Liu, H.; Qin, J. R2YOLOX: A Lightweight Refined Anchor-Free Rotated Detector for Object Detection in Aerial Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5632715. [CrossRef]
24. Peng, G.; Yang, Z.; Wang, S.; Zhou, Y. AMFLW-YOLO: A Lightweight Network for Remote Sensing Image Detection Based on Attention Mechanism and Multiscale Feature Fusion. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4600916. [CrossRef]

25. Guan, R.; Li, Z.; Tu, W.; Wang, J.; Liu, Y.; Li, X.; Tang, C.; Feng, R. Contrastive Multi-View Subspace Clustering of Hyperspectral Images Based on Graph Convolutional Networks. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5510514. [CrossRef]
26. Guan, R.; Li, Z.; Li, T.; Li, X.; Yang, J.; Chen, W. Classification of Heterogeneous Mining Areas Based on ResCapsNet and Gaofen-5 Imagery. *Remote Sens.* **2022**, *14*, 3216. [CrossRef]
27. Xie, S.; Zhou, M.; Wang, C.; Huang, S. CSPPartial-YOLO: A Lightweight YOLO-Based Method for Typical Objects Detection in Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 388–399. [CrossRef]
28. Lv, Q.; Quan, Y.; Sha, M.; Feng, W.; Xing, M. Deep Neural Network-Based Interrupted Sampling Deceptive Jamming Countermeasure Method. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 9073–9085. [CrossRef]
29. Li, Y.; Zhang, S.; Wang, W.-Q. A Lightweight Faster R-CNN for Ship Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 4006105. [CrossRef]
30. Zhang, W.; Liu, Z.; Zhou, S.; Qi, W.; Wu, X.; Zhang, T.; Han, L. LS-YOLO: A Novel Model for Detecting Multi-Scale Landslides with Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 4952–4965. [CrossRef]
31. Xu, Q.; Li, Y.; Shi, Z. LMO-YOLO: A Ship Detection Model for Low-Resolution Optical Satellite Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4117–4131. [CrossRef]
32. Zhao, Z.; Du, J.; Li, C.; Fang, X.; Xiao, Y.; Tang, J. Dense Tiny Object Detection: A Scene Context Guided Approach and a Unified Benchmark. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5606913. [CrossRef]
33. Zhang, L.; Wang, M.; Ding, Y.; Bu, X. MS-FRCNN: A Multi-Scale Faster RCNN Model for Small Target Forest Fire Detection. *Forests* **2023**, *14*, 616. [CrossRef]
34. Pan, J.; Ou, X.; Xu, L. A Collaborative Region Detection and Grading Framework for Forest Fire Smoke Using Weakly Supervised Fine Segmentation and Lightweight Faster-RCNN. *Forests* **2021**, *12*, 768. [CrossRef]
35. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
36. Niu, Z.; Zhong, G.; Yu, H. A Review on the Attention Mechanism of Deep Learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
37. Luo, M.; Xu, L.; Yang, Y.; Cao, M.; Yang, J. Laboratory Flame Smoke Detection Based on an Improved YOLOX Algorithm. *Appl. Sci.* **2022**, *12*, 12876. [CrossRef]
38. Xue, Q.; Lin, H.; Wang, F. FCDM: An Improved Forest Fire Classification and Detection Model Based on YOLOv5. *Forests* **2022**, *13*, 2129. [CrossRef]
39. Chen, G.; Cheng, R.; Lin, X.; Jiao, W.; Bai, D.; Lin, H. LMDFS: A Lightweight Model for Detecting Forest Fire Smoke in UAV Images Based on YOLOv7. *Remote Sens.* **2023**, *15*, 3790. [CrossRef]
40. Wang, G.; Li, H.; Li, P.; Lang, X.; Feng, Y.; Ding, Z.; Xie, S. M4SFWD: A Multi-Faceted Synthetic Dataset for Remote Sensing Forest Wildfires Detection. *Expert. Syst. Appl.* **2024**, *248*, 123489. [CrossRef]
41. Shamsoshoara, A.; Afghah, F.; Razi, A.; Zheng, L.; Fulé, P.Z.; Blasch, E. Aerial Imagery Pile Burn Detection Using Deep Learning: The FLAME Dataset. *Comput. Netw.* **2021**, *193*, 108001. [CrossRef]
42. Wang, Z.; Hua, Z.; Wen, Y.; Zhang, S.; Xu, X.; Song, H. E-YOLO: Recognition of Estrus Cow Based on Improved YOLOv8n Model. *Expert. Syst. Appl.* **2024**, *238*, 122212. [CrossRef]
43. Yi, H.; Liu, B.; Zhao, B.; Liu, E. Small Object Detection Algorithm Based on Improved YOLOv8 for Remote Sensing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *17*, 1734–1747. [CrossRef]
44. Neubeck, A.; Van Gool, L. Efficient Non-Maximum Suppression. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; IEEE: Piscataway, NJ, USA, 2006; Volume 3, pp. 850–855.
45. Tang, Y.; Han, K.; Guo, J.; Xu, C.; Xu, C.; Wang, Y. GhostNetv2: Enhance Cheap Operation with Long-Range Attention. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 9969–9982.
46. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More Features from Cheap Operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1580–1589.
47. Tan, H.; Liu, X.; Yin, B.; Li, X. MHSA-Net: Multihead Self-Attention Network for Occluded Person Re-Identification. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 8210–8224. [CrossRef] [PubMed]
48. Guo, M.-H.; Lu, C.-Z.; Hou, Q.; Liu, Z.; Cheng, M.-M.; Hu, S.-M. Segnext: Rethinking Convolutional Attention Design for Semantic Segmentation. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 1140–1156.
49. Adarsh, P.; Rathi, P.; Kumar, M. YOLO V3-Tiny: Object Detection and Recognition Using One Stage Improved Model. In Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 6–7 March 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 687–694.
50. Jocher, G.; Chaurasia, A.; Stoken, A.; Borovec, J.; Kwon, Y.; Michael, K.; Fang, J.; Wong, C.; Yifu, Z.; Montes, D. Ultralytics/Yolov5: V6. 2-Yolov5 Classification Models, Apple M1, Reproducibility, Clearml and Deci. Ai Integrations. *Zenodo* **2022**. [CrossRef]
51. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In *Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023*; pp. 7464–7475.
52. Girshick, R. Fast R-Cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

53. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision–ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2016; Volume 9905, pp. 21–37. ISBN 978-3-319-46447-3.
54. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv* **2022**, arXiv:2209.02976.

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.