

# Real-Time Early Detection of Forest Fires Using Various YOLO11 Architectures

Sarvesh S Sankop, Anoop U Kadakol, Md Sami A Ghorl,  
Prithvi M Ganiger, Uday Kulkarni and Shashank Hegde

*School of Computer Science and Engineering, KLE Technological University, Hubli, Karnataka, India*

**Keywords:** Forest Fire Detection, YOLO, Deep Learning, Real-Time Detection, Satellite Imagery, Drone Imagery, Transfer Learning, Data Augmentation, Computational Efficiency, Edge Computing.

**Abstract:** Forest fires are an important environmental hazard threatening biodiversity, ecosystems, and human safety. Early and accurate detection is critical for minimizing damage and ensuring timely intervention. This proposed work leverages the YOLO11 architecture to create a robust and efficient forest fire detection framework. The proposed approach achieves superior detection performance under diverse environmental conditions by introducing advanced modules such as C3K2, SPFF, and C2PSA, and leveraging transfer learning. The framework is trained on a heterogeneous dataset combining satellite, drone, and ground-level imagery, capturing a wide spectrum of forest fire scenarios across varying terrains and lighting conditions. The inclusion of data augmentation techniques enhances model generalization to unseen fire patterns. The results indicate that YOLO11-M achieves the best trade-off between precision (85.3%) and recall (81.1%), with a mean average precision (mAP @ 50) of (84.9%), while YOLO11-N offers high computational efficiency for deployment in resource-constrained environments such as edge devices. Furthermore, the integration of real-time detection capabilities enables rapid response to forest fire outbreaks, making this framework a valuable tool for disaster prevention, ecological preservation, and safeguarding human lives.

## 1 INTRODUCTION

Forest fires are among the most destructive environmental disasters, causing widespread damage to biodiversity, wildlife habitats, and human communities. These fires devastate ecosystems that have developed over centuries, pushing endangered species closer to extinction by erasing their natural habitats (Kim and Muminov, 2023; Mohnish et al., 2022; Zhang et al., 2023b). Furthermore, smoke and pollutants released during wildfires lead to climate change, creating a cycle of environmental degradation. Mountainous regions, known for their dense forests and diverse species, are vulnerable due to rugged terrains and weather patterns that allow fires to spread rapidly (Jiao et al., 2019; Zhang et al., 2023a).

Traditional forest fire detection methods, such as manual patrols and satellite monitoring, have limited the damage but are often slow, costly, and ineffective, particularly in remote areas (Ji et al., 2024; Jia et al., 2023). These limitations lead to delayed responses, allowing fires to grow uncontrollably and cause significant losses. With forest fires becoming more fre-

quent and severe due to climate change and human activities, there is an urgent need for advanced, real-time solutions to detect fires early and predict fire-prone areas (Akhlofi et al., 2020; Prakash et al., 2023; Zhang et al., 2022).

The proposed research addresses these challenges by proposing a modern solution that utilizes the YOLO11 (Ali and Zhang, 2024) deep learning framework for real-time forest fire detection and prediction. YOLO11's high accuracy, speed, and real-time processing capabilities make it an ideal candidate for this task (Vasconcelos et al., 2024). Integrated with drone technology, the system can capture high-resolution images of forested areas, detecting early fire signs such as smoke and small flames, even in remote and rugged terrain (Fodor and Conde, 2023). Furthermore, by using meteorological data such as temperature, humidity, and wind speed, the system can predict fire-prone regions and facilitate proactive measures to prevent outbreaks (Shroff, 2023; Zhao et al., 2022a; Zhang et al., 2023b).

A critical component of this study is using satellite images, which provide data for monitoring vast

and difficult-to-reach forest regions. Satellite imagery offers a broad perspective and the ability to capture high-resolution photos across different time frames, which is important for detecting early signs of fire, such as smoke plumes and changes in vegetation. Our dataset, which includes satellite images from various sources, is the foundation for training the YOLO11(Ali and Zhang, 2024) model. The diverse dataset contains multiple environmental conditions, geographical features, and fire events, making it ideal for building a robust fire detection system(Yang et al., 2024; Li et al., 2022).

The paper explains the architecture and innovations of the model, as well as its application to real-time forest fire detection and prediction. Section 2 outlines the background and related works, providing insights into the evolution of YOLO (Wang et al., 2021a) models and their limitations in similar tasks, and discusses YOLO11's Architecture. Section 3 describes the proposed framework, including details about data collection, preprocessing, and training methodologies. Moving on to the Results and Discussion Section 4 evaluates the performance of different YOLO11 variants, emphasizing their suitability for diverse scenarios. Finally, the Conclusion and Future Work Section 5 provides a summary of the overall contributions of the work and proposes directions for future research to enhance the system's scalability, efficiency, and adaptability to diverse real-world scenarios.

## 2 BACKGROUND AND RELATED WORKS

The YOLO series has evolved significantly to address the increasing demands of real-time object detection tasks. YOLOv1 (Zhao et al., 2022b) introduced single-pass object detection for rapid inference but faced challenges with small objects and dense environments. YOLOv2 (Wu and Zhang, 2018) and YOLOv3 (Al-Smadi et al., 2023) brought improvements such as batch normalization, anchor boxes, and multi-scale detection, enhancing accuracy but still struggling to detect subtle features like faint smoke or small flames. Further advancements, including YOLOv4 (Zhao et al., 2022c; Wang et al., 2021b) and YOLOv5, focused on computational efficiency and adaptability to edge devices. The subsequent iterations, YOLOv6 through YOLOv10 (Talaat and ZainEldin, 2023; Huynh et al., 2024), refined these aspects while addressing false positives and overlapping detections. However, processing high-resolution satellite imagery remained a challenge, highlighting

the need for further innovation in forest fire detection tasks.

YOLO11 addresses these challenges with architectural enhancements like the C3K2 block, SPFF module, and C2PSA block. These components improve spatial feature extraction and multi-scale feature aggregation, ensuring precise detection even under complex conditions. With superior speed and accuracy, YOLO11 bridges the gap between real-time performance and scalability, outperforming state-of-the-art methods such as Transformer-based object detectors which makes YOLO11 a robust and reliable solution for detecting forest fires across diverse environmental conditions and image resolutions.

### 2.1 Architecture of YOLO11

The YOLO11 architecture comprises three main components: Backbone, Neck, and Head, which work together to achieve high accuracy, scalability, and efficiency in various object detection tasks. Figure 1 shows the overall architecture of YOLO11, detailing the flow from the Backbone to the Head.

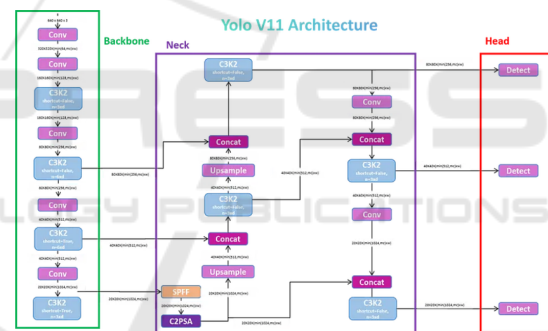


Figure 1: An overview of the YOLO11 architecture, highlighting its backbone, neck, and head components.

#### 2.1.1 Backbone

The Backbone extracts spatial features from input images using convolutional layers. It applies convolution, activation, and downsampling operations to identify patterns such as edges and textures. This process can be mathematically represented in Equation 1, explaining convolution operation:

$$\text{Output}_{\text{conv}} = \text{ReLU}(\text{Conv}(\mathbf{X}, \mathbf{W}) + \mathbf{b}) \quad (1)$$

Here,  $\mathbf{X}$  represents the input,  $\mathbf{W}$  is the filter kernel, and  $\mathbf{b}$  is the bias term. Equation 1 explains how combining convolution and ReLU activation extracts essential features. Residual connections in the C3 blocks, shown in Figure 1, improve gradient flow and enable multi-resolution feature extraction.

### 2.1.2 Neck

The Neck is an intermediary between the Backbone and the Head, aggregating multiresolution features to enable effective object detection across various scales. Figure 1 highlights the role of the Neck in fusing features for improved detection.

It uses Spatial Pyramid Pooling-Fast (SPPF) to capture contextual information across multiple scales, as shown in the middle section of Figure 1. This method allows the model to detect objects even in complex scenarios effectively. Cross-scale connections further enhance feature integration, preserving fine-grained details critical for detecting small or overlapping objects.

### 2.1.3 Head

The Head generates the final predictions, including bounding box coordinates, class probabilities, and objectness scores. Bounding box predictions  $(x, y, w, h)$  are offsets relative to predefined anchor boxes. Equation 2 explains how class probabilities for each bounding box are computed using the softmax function:

$$P_{\text{class}} = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (2)$$

Here,  $z_i$  represents the score for class  $i$ , and the softmax ensures the probabilities sum to 1, making predictions interpretable.

Equation 3 explains the objectness score, which uses the sigmoid function to determine whether a bounding box likely contains an object:

$$S_{\text{object}} = \sigma(z) \quad (3)$$

The Head operates across multiple scales to detect objects of varying sizes, as depicted in Figure 1. Non-Maximum Suppression (NMS) eliminates redundant detections, retaining the most confident predictions. This design ensures precision and efficiency, making YOLO11 suitable for real-time applications.

## 3 PROPOSED METHODOLOGY

The proposed framework Figure 3 aims to utilize the latest advances in object detection techniques to accurately and efficiently identify forest fires using satellite and drone imagery. The framework leverages YOLO11's architectural enhancements, such as the C3K2 block, SPPF module, and C2PSA block, to ensure superior performance in real-time detection tasks. The major steps of the proposed work are given below:

### 3.1 Dataset and Preprocessing

The development of an efficient forest fire detection system began with the creation of a custom dataset comprising 5,514 images collected from satellite and drone imagery. The dataset captured variations in fire intensities, lighting conditions, and obstructions such as smoke, reflecting real-world complexities as shown in Figure 2.



Figure 2: Sample Images from our Dataset which includes Satellite and Drone Images.

To facilitate training and evaluation, the dataset was divided into three subsets: 80% (4,411 images) for training, 10% (551 images) for validation, and 10% (552 images) for testing.

This division ensured a large training set for model learning while providing balanced validation and test sets for unbiased performance benchmarks. Each subset was carefully composed to include diverse conditions such as fire intensities, times of day, and environmental obstructions.

The data set collected was preprocessed to ensure compatibility with YOLO11 models and improve training quality. Images were resized to 640×640 pixels to standardize dimensions and optimize computational efficiency. Annotation of flames, smoke, and scorched areas was performed using tools like LabelImg and Roboflow, providing accurate bounding boxes for supervised learning.

To enhance model generalization, data augmentation techniques such as flips, brightness adjustments, and rotations were applied, simulating various environmental and perspective conditions. These preprocessing steps ensured the dataset's robustness and representativeness of real-world forest fire scenarios.

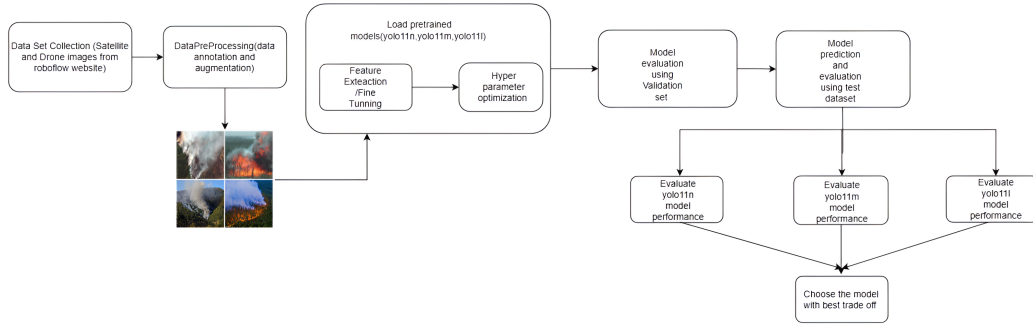


Figure 3: Proposed Workflow of our methodology for forest fire detection, including dataset collection, preprocessing, feature extraction, hyperparameter optimization, model evaluation, and selection of the best-performing YOLO model.

### 3.2 Feature Extraction and Fine-Tuning

Feature extraction was performed using pre-trained YOLO models initially trained on large-scale datasets. These models extracted critical features such as texture, shape, and edge details relevant to forest fire detection. The initial layers of the YOLO models were frozen to preserve generic feature representations, while the later layers were fine-tuned using the forest fire dataset to adapt to domain-specific patterns like smoke and fire signatures. Key parameters, such as learning rate, were adjusted during this process to improve detection accuracy.

The feature extraction process is represented as:

$$F_{\text{forest}} = f(W_{\text{pretrained}}, X_{\text{input}}) \quad (4)$$

Here,  $F_{\text{forest}}$  denotes the feature map for forest fire detection,  $W_{\text{pretrained}}$  refers to the pre-trained YOLO weights, and  $X_{\text{input}}$  represents the input image. Equation 4 explains how features are extracted from pre-trained YOLO models and adapted to the forest fire dataset.

### 3.3 Hyperparameter Optimization

Key hyperparameters, including learning rate, batch size, and number of epochs, were tuned to enhance model performance. Both grid search and random search techniques were employed to identify optimal configurations. Grid search explored predefined parameter ranges, while random search sampled values within specified intervals. This hybrid approach ensured efficient and effective optimization.

The optimization process can be expressed as:

$$H^* = \arg \min_H \mathcal{L}(H; D_{\text{train}}) \quad (5)$$

Here,  $H$  represents the set of hyperparameters,  $\mathcal{L}$  is the loss function, and  $D_{\text{train}}$  is the training dataset.

Equation 5 illustrates the optimization process used to fine-tune hyperparameters for effective training.

### 3.4 Model Training

The YOLO11 loss function optimizes three aspects: accurate classification, reliable confidence, and precise localization. The total loss is expressed as:

$$L = \lambda_{\text{cls}} \cdot L_{\text{cls}} + \lambda_{\text{obj}} \cdot L_{\text{obj}} + \lambda_{\text{box}} \cdot L_{\text{box}} \quad (6)$$

Equation 6 explains how the total loss function combines classification, confidence, and localization losses for training YOLO11 models.

- Class Loss ( $L_{\text{cls}}$ ) ensures accurate object categorization, penalizing incorrect classifications:

$$L_{\text{cls}} = - \sum_{i=1}^C p_i \log(\hat{p}_i) \quad (7)$$

Equation 7 explains how classification accuracy is calculated by comparing predicted and true probabilities.

- Object Loss ( $L_{\text{obj}}$ ) evaluates prediction confidence using Binary Cross-Entropy:

$$L_{\text{obj}} = - [y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})] \quad (8)$$

Equation 8 explains how objectness confidence is penalized based on prediction and ground truth.

- Box Loss ( $L_{\text{box}}$ ) refines bounding box alignment using Complete Intersection over Union (CIoU):

$$L_{\text{box}} = 1 - \text{IoU} + \frac{\rho^2(b, b_{\text{gt}})}{c^2} + \alpha_v \quad (9)$$

Equation 9 explains how IoU improves bounding box accuracy by minimizing distance and aspect ratio differences.

This design ensures that YOLO11 is effective in classifying and localizing fire indicators like smoke and flames under diverse conditions while maintaining efficiency for real-time applications.



## 4 RESULTS AND DISCUSSION

This section presents the results and analysis of the YOLO11-N, YOLO11-M, and YOLO11-L models applied to the forest fire detection task. The evaluation is based on three primary performance metrics: mean Average Precision (mAP@50), Precision, and Recall. A comparative discussion highlights the strengths, weaknesses, and trade-offs of each model in terms of detection accuracy, computational efficiency, and suitability for real-time deployment in diverse scenarios.

### 4.1 YOLO11-N

The YOLO11-N variant achieved a mean Average Precision (mAP@50) of 81.8%, with Precision at 84.2% and Recall at 79.2%. This model is lightweight and computationally efficient, making it suitable for deployment on edge devices. The rapid convergence of loss components, such as Box Loss and Object Loss, indicates its effectiveness in learning core object detection tasks within the first 150 epochs.

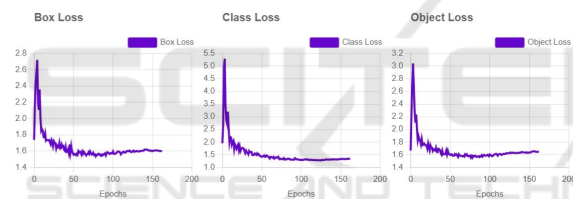


Figure 4: Training loss curves for the YOLO11-N variant, showing the convergence of Box Loss, Class Loss, and Object Loss over 150 epochs.

However, its slightly lower recall compared to YOLO11-M suggests it may miss some smaller or less distinct fire indicators. Figure 4 represents the loss curves for the YOLO11-N model, showcasing its fast convergence across all training components.

### 4.2 YOLO11-M

Among the three variants, YOLO11-M demonstrated the highest performance, with a mean Average Precision (mAP@50) of 84.9%, Precision of 85.3%, and Recall of 81.1%. This balance between Precision and Recall makes it the most suitable for applications requiring high detection accuracy and sensitivity. The smooth loss convergence and consistent generalization indicate that YOLO11-M effectively captures both spatial and contextual features critical for forest fire detection. The Figure 5 illustrates the training loss curves for YOLO11-M. It emphasizes the smooth and stable convergence of the total loss and

its components. Its performance highlights its robustness in detecting subtle fire indicators across varying environmental conditions.

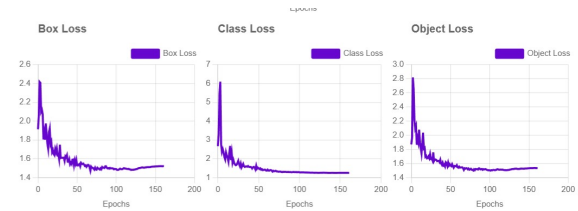


Figure 5: Loss curves for YOLO11-M during training, demonstrating smooth and stable convergence across all components.

### 4.3 YOLO11-L

YOLO11-L prioritized Precision, achieving a score of 89.1%, but with a lower Recall of 77.4% and mAP@50 of 80.4%. Its high Precision minimizes false positives, making it ideal for applications where accuracy is critical, such as disaster management or high-security environments. However, its reduced Recall suggests limitations in detecting smaller or partially occluded fire indicators. The model's com-



Figure 6: Training loss curves for the YOLO11-L variant.

putational demands also make it less suitable for real-time applications on resource-constrained devices. Figure 6 highlights the YOLO11-L training loss curves, reflecting its prioritization of Precision over Recall.

#### 4.3.1 Comparative Analysis

The comparative performance analysis reveals that YOLO11-M strikes the best balance between accuracy, sensitivity, and computational efficiency. While YOLO11-N is optimal for lightweight applications, YOLO11-L is preferable for precision-critical tasks. The diverse strengths of these models make them adaptable for various deployment scenarios, from real-time edge computing to high-accuracy monitoring systems. Table 1 illustrates the comparative performance of the three YOLO11 variants in terms of mAP@50, Precision, and Recall.

Table 1: Performance comparison of YOLO11-N, YOLO11-M, and YOLO11-L models based on mAP@50, Precision, and Recall metrics.

Model	mAP@50(%)	Precision(%)	Recall(%)
YOLO11-N	81.8	84.2	79.2
YOLO11-M	84.9	85.3	81.1
YOLO11-L	80.4	89.1	77.4

4.3.2 Visual Analysis



Figure 7: Sample Output of our YOLO11 model identifying smoke and flames in diverse forest conditions. The bounding boxes display the model’s capability to detect small fires and faint smoke effectively.

Qualitative results, including bounding box visualizations, illustrate the models’ capabilities in accurately identifying flames, smoke, and scorched areas under diverse environmental conditions. These visualizations confirm the adaptability of the YOLO11 framework for real-world forest fire detection tasks, even in challenging scenarios with occlusions, dense vegetation, and varying lighting conditions Figure 7 provides qualitative visualizations of the YOLO11 framework, showing its ability to detect flames and smoke in challenging environmental conditions.

5 CONCLUSION AND FUTURE SCOPE

This study proposes an advanced framework for real-time forest fire detection using the innovative

YOLO11 architecture. With components like the C3K2 block, SPFF module, and C2PSA block, YOLO11 excels in detecting early fire indicators such as smoke and flames. Among its variants, YOLO11-M achieves the best balance between precision and recall, while YOLO11-N is ideal for resource-constrained environments, and YOLO11-L minimizes false positives for critical applications. Tested on diverse satellite and drone imagery, the framework demonstrates robust performance, making YOLO11 a reliable tool for early fire detection and management.

Future work includes expanding the dataset to cover more diverse conditions and integrating multi-modal data like thermal imaging and meteorological inputs for enhanced accuracy. Deploying YOLO11 on lightweight edge devices such as drones and IoT sensors can enable real-time monitoring in remote areas. Additionally, combining YOLO11 with fire spread modeling could offer a comprehensive solution for proactive fire management, reducing wildfire impacts globally.

REFERENCES

Akhoulfi, M. A., Castro, N. A., and Couturier, A. (2020). Unmanned aerial systems for wildland and forest fires. *arXiv preprint arXiv:2004.13883*.

Al-Smadi, Y., Alauthman, M., Al-Qerem, A., Aldweesh, A., Quaddoura, R., Aburub, F., Mansour, K., and Alhmiedat, T. (2023). Early wildfire smoke detection using different yolo models. *Machines*, 11(2):246.

Ali, M. L. and Zhang, Z. (2024). The yolo framework: A comprehensive review of evolution, applications, and benchmarks in object detection. *Computers*, 13(12):336.

Fodor, G. and Conde, M. V. (2023). Rapid deforestation and burned area detection using deep multi-modal learning on satellite imagery. *arXiv preprint arXiv:2307.04916*.

Huynh, T. T., Nguyen, H. T., and Phu, D. T. (2024). Enhancing fire detection performance based on fine-tuned yolov10. *Computers, Materials & Continua*, 81(2).

Ji, C., Yang, H., Li, X., Pei, X., Li, M., Yuan, H., Cao, Y., Chen, B., Qu, S., Zhang, N., et al. (2024). Forest wildfire risk assessment of anning river valley in sichuan province based on driving factors with multi-source data. *Forests*, 15(9):1523.

Jia, X., Wang, Y., and Chen, T. (2023). Forest fire detection and recognition using yolov8 algorithms from uavs images. In *2023 IEEE 5th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, pages 646–651.

Jiao, Z., Zhang, Y., Xin, J., Mu, L., Yi, Y., Liu, H., and Liu, D. (2019). A deep learning based forest fire detection

- approach using uav and yolov3. In *2019 1st International conference on industrial artificial intelligence (IAI)*, pages 1–5. IEEE.
- Kim, S.-Y. and Muminov, A. (2023). Forest fire smoke detection based on deep learning approaches and unmanned aerial vehicle images. *Sensors*, 23(12):5702.
- Li, Y., Shen, Z., Li, J., and Xu, Z. (2022). A deep learning method based on srn-yolo for forest fire detection. In *2022 5th International Symposium on Autonomous Systems (ISAS)*, pages 1–6.
- Mohnish, S., Akshay, K., Pavithra, P., Ezhilarasi, S., et al. (2022). Deep learning based forest fire detection and alert system. In *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, pages 1–5. IEEE.
- Prakash, M., Neelakandan, S., Tamilselvi, M., Velmurugan, S., Baghavathi Priya, S., and Ofori Martinson, E. (2023). Deep learning-based wildfire image detection and classification systems for controlling biomass. *International Journal of Intelligent Systems*, 2023(1):7939516.
- Shroff, P. (2023). Ai-based wildfire prevention, detection and suppression system. *arXiv preprint arXiv:2312.06990*.
- Talaat, F. M. and ZainEldin, H. (2023). An improved fire detection approach based on yolo-v8 for smart cities. *Neural Computing and Applications*, 35(28):20939–20954.
- Vasconcelos, R. N., Franca Rocha, W. J., Costa, D. P., Duverger, S. G., Santana, M. M. d., Cambui, E. C., Ferreira-Ferreira, J., Oliveira, M., Barbosa, L. d. S., and Cordeiro, C. L. (2024). Fire detection with deep learning: A comprehensive review. *Land*, 13(10):1696.
- Wang, S., Chen, T., Lv, X., Zhao, J., Zou, X., Zhao, X., Xiao, M., and Wei, H. (2021a). Forest fire detection based on lightweight yolo. In *2021 33rd Chinese Control and Decision Conference (CCDC)*, pages 1560–1565. IEEE.
- Wang, S., Chen, T., Lv, X., Zhao, J., Zou, X., Zhao, X., Xiao, M., and Wei, H. (2021b). Forest fire detection based on lightweight yolo. In *2021 33rd Chinese Control and Decision Conference (CCDC)*, pages 1560–1565.
- Wu, S. and Zhang, L. (2018). Using popular object detection methods for real time forest fire detection. In *2018 11th International symposium on computational intelligence and design (ISCID)*, volume 1, pages 280–284. IEEE.
- Yang, S., Huang, Q., and Yu, M. (2024). Advancements in remote sensing for active fire detection: a review of datasets and methods. *Science of the total environment*, page 173273.
- Zhang, L., Li, J., and Zhang, F. (2023a). An efficient forest fire target detection model based on improved yolov5. *Fire*, 6(8).
- Zhang, L., Wang, M., Ding, Y., and Bu, X. (2023b). Ms-frcnn: A multi-scale faster rcnn model for small target forest fire detection. *Forests*, 14(3):616.
- Zhang, Y., Chen, S., Wang, W., Zhang, W., and Zhang, L. (2022). Pyramid attention based early forest fire detection using uav imagery. In *Journal of Physics: Conference Series*, volume 2363, page 012021. IOP Publishing.
- Zhao, L., Zhi, L., Zhao, C., and Zheng, W. (2022a). Fire-yolo: A small target object detection method for fire inspection. *Sustainability*, 14(9).
- Zhao, L., Zhi, L., Zhao, C., and Zheng, W. (2022b). Fire-yolo: a small target object detection method for fire inspection. *Sustainability*, 14(9):4930.
- Zhao, S., Liu, B., Chi, Z., Li, T., and Li, S. (2022c). Characteristics based fire detection system under the effect of electric fields with improved yolo-v4 and vibe. *IEEE Access*, 10:81899–81909.