ISSN: 2508-7894
KJAI website: http://accesson.kr/kjai
doi: http://dx.doi.org/10.24225/kjai.2024.12.4.27

Learnable Sobel Filter and Attention-based Deep Learning Framework for Early Forest Fire Detection

Sehun KIM[1], Kyeongseok JANG[2], Dongwoo LEE[3], Seungwon CHO[4], Seunghyun LEE[5], Kwangchul SON[6*]

Received: November 04, 2024. Revised: November 19, 2024. Accepted: December 05, 2024.

Abstract

Various techniques are being researched to effectively detect forest fires. Among them, techniques using object detection models can monitor forest fires over wide areas 24 hours a day. However, detecting forest fires early with traditional object detection models is a very challenging task. While they show decent accuracy for thick smoke and large fires, they show low accuracy for faint smoke and small fires, and frequently generate false positives for lights that are like fires. In this paper, to solve these problems, we focus on leveraging local characteristics such as contours and textures of fire and smoke, which are crucial for accurate detection. Based on this approach, we propose EDAM (Edge driven Attention Module) that performs enhancement by richly utilizing contour and texture information of fire and smoke. EDAM extracts important edge information to generate feature maps with emphasized contour and texture information, and based on this map, performs Attention Mechanism to emphasize key characteristics of smoke and fire. Through this mechanism, the overall model performance was improved, with AP_S increasing from 0.154 to 0.204 and $AP_{0.5}$ from 0.779 to 0.784, resulting in a significant improvement in AP_S value to 32.47%. In practice, the model applying this technique showed excellent inference speed while greatly improving detection performance for small objects compared to existing models and reduced false positive rates for building and street light illumination in nighttime environments that are easily mistaken for fire.

Keywords : Artificial Intelligence, Real-time Detection, Feature Extraction, Attention Mechanism, Early Wildfire Detection

Major Classification Code : Artificial Intelligence, Deep Learning, Object Detection, wildFire Detection, etc

1. Introduction

As greenhouse gases increase due to frequent use of fossil fuels and various human activities such as internal

* This research was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency(KOCCA) grant funded by the Ministry of Culture, Sports and Tourism(MCST) in 2024(Project Name: 3D holographic infotainment system design R&D professional human resources, Project Number: RS-2024-00401213, Contribution Rate: 100%)

1 First Author. student, Defense Acquisition Program, Kwangwoon University, Republic of Korea. Email: rlatpgns0417@kw.ac.kr
2 Second Author. student, Department of Plasma Bio Display, Kwangwoon University, Republic of Korea. Email: ksjang1234@kw.ac.kr
3 Third Author. student, Department of Plasma Bio Display, Kwangwoon University, Republic of Korea. Email: led0121@kw.ac.kr
4 Fourth Author, Student, Department of Holographic 3D Contents, Kwangwoon University, Republic of Korea, Email: cswhm22@hanmail.net
5 Fifth Author, Professor, Department of Ingenisum College Liberal Arts, Kwangwoon University, Republic of Korea, Email: shlee@kw.ac.kr
6 Corresponding Author, Professor, Department of Electronics and Communications Engineering, Kwangwoon University, Republic of Korea, Email: kcson@kw.ac.kr

combustion engine vehicles and thermal power generation, the global warming issue is intensifying with an accelerated rise in Earth's average temperature. This is affecting us directly as a climate crisis beyond climate change. The climate crisis appears in the form of various climate disasters such as droughts, heat waves, forest fires, and typhoons, causing significant damage to various regions, and their scale and frequency are gradually increasing year by year[1]. Among these, what makes forest fires particularly dangerous is that they spread rapidly at very high speeds when conditions of dry weather and strong winds are present. The recent 2024 wildfire in California, United States, expanded into a massive fire that destroyed 8,580 hectares over three days due to hot, dry weather and strong winds, causing enormous human casualties and property damage, while the 2022 Uljin-Samcheok wildfire in South Korea, which started from a small spark, rapidly spread due to drought and strong winds, destroying 20,923 hectares and becoming the largest wildfire in Korean history. According to the Korea Forest Service (KFS), the golden time for early suppression of forest fires is defined as within 50 minutes from the report receipt to the dispatch of forest helicopters. If this golden time is exceeded, the fire can spread rapidly, leading to large-scale forest fires that can destroy tens of thousands of hectares in an instant. Such delayed response makes firefighting operations more difficult and makes it harder to secure evacuation time, ultimately resulting in significant property damage and casualties (Scott et al., 2013). Countries affected by forest fires are establishing fire detection systems in various ways for early suppression before they develop into large-scale forest fires. There are two main methods for monitoring forest fires, using IoT sensors and using Computer Vision technology. The method using IoT sensors utilizes various sensors such as gas sensors, flame detection sensors, and humidity sensors to detect toxic gases and light produced during combustion. These sensors' data are monitored, and when data exceeding threshold values is received, it is determined as a fire (Yu et al., 2005). With the advance of deep learning technology, the research is being conducted on improving fire detection accuracy and predicting fire scale and damaged areas using IoT sensors and Deep Neural Networks (DNN) (Zope et al., 2020). Additionally, by applying Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM), which are specialized for time-series data, it becomes possible to more accurately detect and predict fire patterns by learning data changes and patterns over time (Liu et al., 2023). However, despite high initial installation and maintenance costs, early fire detection is only possible in limited areas, making it impractical for use in vast mountains and forests. Recently, to overcome these limitations, the research is being conducted on detecting fires in wide areas using Computer Vision technology. The Computer Vision method involves detecting smoke and fire using artificial intelligence technologies such as Semantic Segmentation (Al-Dabbagh & Ilyas, 2023) or Object Detection (Jindal et al., 2021) in video data received from drones or CCTV.

The key to early detection of forest fires is not the accurate shape of smoke or fire, but rather quickly discovering the location where a fire has occurred in real-time. Therefore, object detection, which quickly performs object classification and localization by marking bounding boxes around objects, is mainly used rather than semantic segmentation, which segments and classifies objects in image data at the pixel level. This technology has the potential to minimize forest fire damage by monitoring forest fires 24 hours a day with low maintenance costs and enabling rapid response in the early stages of fires. However, existing object detection models still struggle to detect small and faint white smoke, which is one of the early symptoms of forest fires.

In this paper, to overcome the limitations of low accuracy and false detection of faint smoke and small fires in the early stages of forest fires, we propose an Edge Driven Attention Module (EDAM) that utilizes texture and contour information of smoke and small fires, thereby improving detection accuracy of white smoke and small fires in the early stages of forest fires while reducing false detection frequency.

2. Related Works

2.1. Dataset

Existing publicly available datasets predominantly contain wildfire images taken at close range or images of already developed large-scale wildfires. Since models trained on such datasets are not suitable for early forest fire detection tasks, we collected datasets with early fire images from drones and CCTV. To build the early wildfire detection dataset proposed in this paper, we collected and selected images from Roboflow's Forest Fire and Smoke and Wild Fire Smoke Dataset to construct a dataset suitable for early wildfire detection.

(a) (b)

Figure 1: Unsuitable public dataset images: (a) directly photographed fire, (b) large forest fire

Figure 1 shows examples of datasets that are unsuitable for the purpose of this paper. The publicly available dataset contains a large number of fire images directly taken by people, as shown in (a), and photos taken after fires have already developed into large forest fires, as shown in (b).

(a) (b)

Figure2 : Suitable public dataset images: (a) white smoke captured by drone, (b) fire captured by CCTV

Also, Figure 2 shows examples of datasets suitable for the purpose. (a) shows an image of white smoke in the early stages of a forest fire captured in a drone environment, and (b) shows a nighttime forest fire image captured in a CCTV environment. The dataset consists of 23,777 images in two classes: fire and smoke.

2.2. YOLOv7

You only look once (YOLO) is one of the models used for tasks requiring real-time detection with fast inference speed, operating as a One Stage Object Detection model. Due to continuous development, it shows improved accuracy, lightweight design, and faster inference speed, making it one of the widely used models in the field of Object Detection (Redmon, 2016). In 2022, Chien-Yao Wang proposed Yolov7, which enhanced model performance by applying various techniques in network architecture, training, and testing methods (Wang et al., 2023).

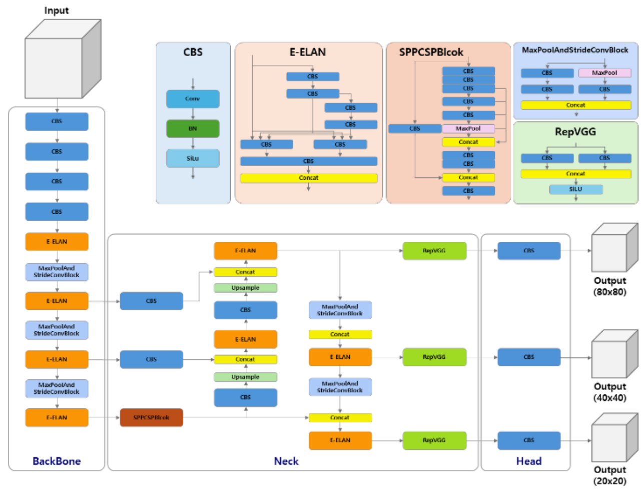

Figure3 : Network Architecture of YOLOv7

Figure 3 shows the overall network architecture of YOLOv7. YOLOv7 incorporates the following techniques. The E-ELAN (Extended Efficient Layer Aggregation Network) was introduced, which extends the existing ELAN (Efficient Layer Aggregation Network) structure (Wang et al., 2022). This structure modifies only the architecture of computational blocks while minimizing changes to the transition layer architecture, continuously improving the network's learning ability without destroying the original gradient path. RepConv is applied between the Neck and Head, maintaining learning capability with a complex structure during training while transforming into a simple structure during inference, significantly improving inference speed. A new scaling technique called Model scaling for concatenation-based models was introduced. This scaling technique calculates output channel changes of blocks and performs scaling on layers, enabling scaling while maintaining both the properties the model had before scaling and optimal structure. The architecture's advanced gradient propagation and efficient scaling mechanisms make it particularly effective for complex detection tasks.

Finally, Trainable bag-of-freebies was introduced to improve object detection accuracy without increasing inference time. These techniques provide the advantage of performing real-time object detection with decent accuracy even in edge computing environments with limited resources, such as CCTV or drones.

2.3. Attention Mechanism

Attention mechanism emphasizes local contextual information by assigning higher weights to important features and lower weights to unnecessary features in input data. The attention mechanism was proposed to address the problem of initial input data being diluted in RNN (Recurrent Neural Network) which performs tasks iteratively (Vaswani, 2017). Subsequently, the attention mechanism expanded from NLP (Natural Language Processing) to computer vision, showing excellent performance improvements in various image processing tasks (Wang et al., 2017). Feature extraction is particularly important in CNN (Convolutional Neural Network)-based computer vision, and accurately capturing features of complex scenes or objects of varying sizes is a crucial factor determining model performance. The attention mechanism further enhances CNN's feature extraction capability by emphasizing important regions while suppressing less important ones.

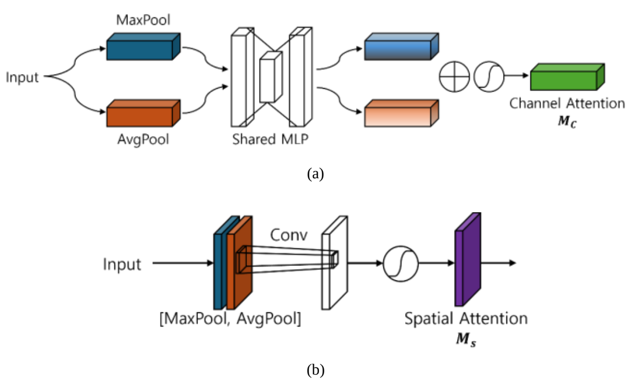

Figure4 : Attention Mechanism Architecture (a)Channel Attention (b)Spatial Attention

Figure 4 shows the structure of Channel Attention and Spatial Attention, which are the main implementation methods of the Attention Mechanism (Fu et al., 2017). Figure 4(a) shows the structure of Channel Attention, which performs MaxPool and AvgPool operations in parallel on the input features, learns channel correlations through Shared MLP, and integrates them to generate a channel attention map (M_c). Figure 4(b) shows the structure of Spatial Attention, which performs MaxPool and AvgPool operations on input features and learns the importance of spatial positions through a Convolution Layer to generate a spatial attention map (M_s). Each Attention can be used independently, but when used in a combined form, it can effectively integrate channel correlations and spatial importance for more comprehensive feature representation (Woo et al., 2018; Dong et al., 2023). This approach significantly improves model accuracy through multidimensional feature emphasis.

3. Proposed Method

Smoke and fire from wildfires have irregular shapes that change according to wind and location. Due to these characteristics, strong feature representation capabilities of the model are required for accurate detection of smoke and fire. To improve the accuracy of early wildfire detection, this paper proposes a new Edge Driven Attention Module (EDAM) that combines Adaptive Edge Enhanced Module (AEEM), which marks important areas using texture and contour information, with Channel Attention and Spatial Attention.

The YOLO model, which is mainly used in wildfire detection systems, shows fast speed capable of real-time processing, but low accuracy has been raised as a drawback. This limitation becomes an even more challenging issue for faint white smoke and small fires in the early stages of wildfires. To address these problems, we propose a framework that applies EDAM to YOLOv7.

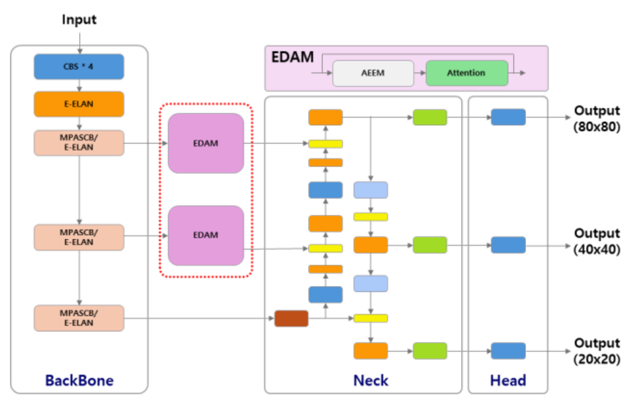

Figure5 : YOLOv7 Architecture with EDAM

Figure 5 shows the proposed network structure based on the existing YOLOv7-L. It is a network that applies the Edge Driven Attention Module to enhance the feature representation of smoke and flame texture characteristics and contour information using edge information in the existing network. EDAM has a modularized structure that can be applied to various network architectures. In this paper, EDAM is placed between the Backbone and Neck of the YOLOv7-L model to enhance the expressiveness of Multi Scale Feature Maps extracted from the Backbone.

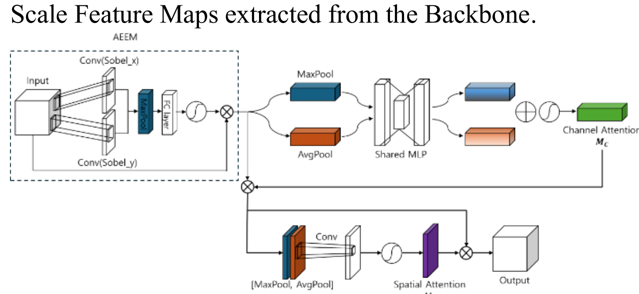

Figure6 : Architecture of EDAM

Figure6 shows the architecture of EDAM that combines AEEM structure with Channel Attention and Spatial Attention. EDAM is a module where AEEM, Channel Attention, and Spatial Attention are applied sequentially. AEEM selectively captures local features such as textures and contours of fire and smoke. This local Edge information is utilized in calculating channel importance during the Channel Attention operation, selectively enhancing channels that well represent characteristic patterns of fire and smoke. Subsequently, through Spatial Attention, important locations in the Attention Map are emphasized to highlight areas containing fire and smoke features. This process improves early wildfire detection capability by considering both local edge features and global information.

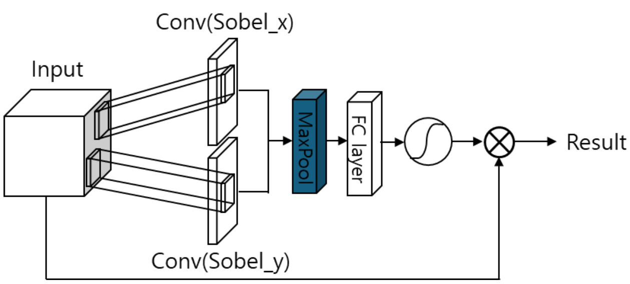

Figure7 : Architecture of AEEM

Figure7 shows the architecture of AEEM, one of the main branches of EDAM. AEEM is an Attention map generation module that utilizes Sobel Convolution from Feature maps. In the edge extraction process using Sobel, learnable parameters are applied to perform adaptive edge extraction. The Sobel filter calculates edges in X and Y directions in the form of (G_x, G_y).

$$Sobel_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix},\ Sobel_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (1)$$

$$Edge = \sqrt{(Edge_x)^2 + \left(Edge_y\right)^2 + \epsilon} \quad (2)$$

Global pooling is performed on the Sobel Convolution results to simplify information, channel importance is learned through a Fully Connected Layer (FC), and finally, edge maps are generated through normalization using a sigmoid function. AEEM acquires the ability to express edges weakly in areas considered less important or noise, and strongly in important areas, performing adaptive edge enhancement to extract texture and contour information of fire and smoke that show diverse and irregular patterns.

4. Result

In this paper, for model performance evaluation, we performed training with 21,287 wildfire images and testing with 415 images that we independently constructed. The dataset consists of scenes containing small fires in early stages or white smoke, captured by drones or CCTV from various locations. For the training strategy, we used Stochastic Gradient Descent (SGD) as the optimizer and Sigmoid Linear Unit (SiLU) as the activation function. The batch size was set to 16, and training was conducted for 100 epochs with a learning rate of $1 \times 10^{(-2)}$. The proposed method and all models are implemented in pytorch (Paszke et al., 2019).

Figure 8-9 shows the comparison of object detection performance between the existing YOLOv7 model and the model with EDAM applied on early-stage wildfire images. Each result compares the performance by showing (1) Ground Truth, (2) detection results of the existing YOLOv7, and (3) detection results of the proposed model with EDAM applied. In fire detection, the model with EDAM accurately identified more fire locations.

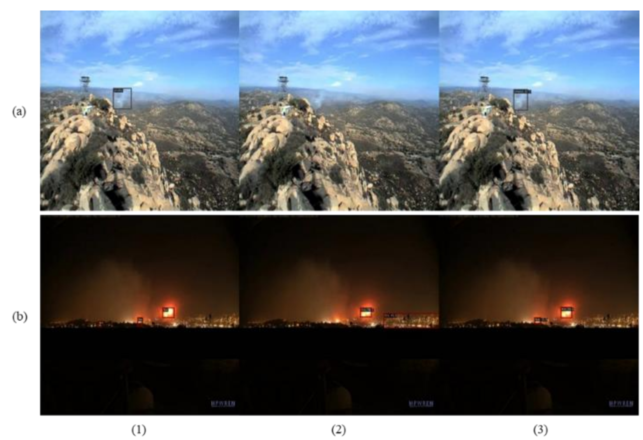

Figure8 : Forest fire images captured by CCTV (a) white smoke image, (b) nighttime fire image, (1) ground truth image, (2) YOLOv7 result, (3) Proposed EDAM result

Figure8 shows the detection results for the early stages of a forest fire captured by CCTV. While (a-2) the existing YOLOv7 model failed to detect faint smoke, (a-3) the model with EDAM successfully detected it. Additionally, (b-2) the existing YOLOv7 model failed to detect small fires and produced false positives for lights, but (b-3) the model with EDAM successfully detected small fires without false positives, showing more reliable detection results.

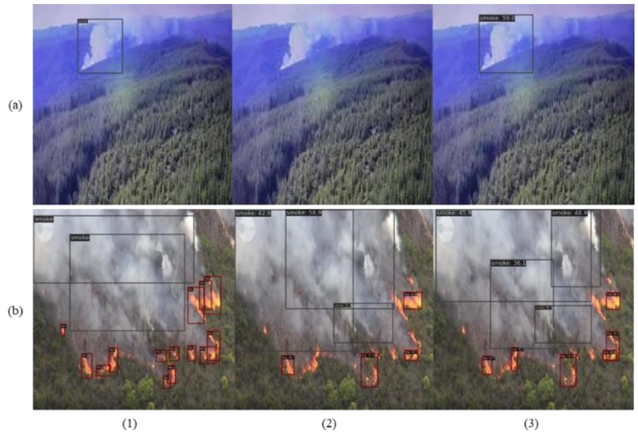

Figure9 : Forest fire images captured by drone (a) white smoke image, (b) fire image, (1) ground truth image, (2) YOLOv7 result, (3) Proposed EDAM result

Figure 9 shows the detection results of forest fires captured by drones. (a-2) The existing YOLOv7 model failed to detect white smoke in the early stages of forest fires, while (a-3) the model with EDAM successfully detected the smoke. (b-2) YOLOv7 only detected some fires, while (b-3) the model with EDAM successfully detected more fire points.

Table 1: Dataset Test Results

Model	$AP_{0.5}$	$AP_{0.75}$	AP_S	AP_M	AP_L	Latency(ms)
YOLOv7	0.779	0.586	0.154	0.408	0.724	0.0214
Proposed Model	0.784	0.604	0.204	0.405	0.729	0.0246

Table 1 compares the wildfire detection results between the model with EDAM applied and the existing YOLOv7 model. The model with EDAM showed overall improved performance compared to the base model, with increased values in $AP_{0.5}$, $AP_{0.75}$, AP_S, and AP_L. Notably, in terms of object size-specific performance, the AP_S value, which indicates small object detection performance, showed a significant improvement of 32.47%, increasing from 0.154 in the base model to 0.204 in the EDAM-applied model. This improvement in small object detection performance suggests enhanced ability to detect fires and smoke in the early stages of wildfires, contributing to securing initial response time for fire suppression. The application of EDAM resulted in a slight increase in inference time due to additional computations required.

5. Conclusions

In this paper, we proposed the AEEM, which performs Sobel operation and emphasizes important texture and contour information through FC Layer and sigmoid operations to enable real-time detection while improving accuracy and reducing false positives. The EDAM, which combines the proposed AEEM with Attention Mechanism, was applied to low-level Feature Maps, enabling efficient object detection for early forest fire symptoms by utilizing object edge information to richly leverage object shape, texture, and contour information. Experimental results showed that YOLOv7 with EDAM more accurately captured early forest fire symptoms in CCTV and drone environments compared to existing models with minimal additional computation. Notably, the model with EDAM successfully detected faint smoke that the existing YOLOv7 could not detect and reduced the frequency of false positives for nighttime lights that could be mistaken for fires. While the differences in AP_M and AP_L values were minimal, the AP_S value significantly improved from 0.154 to 0.204, showing a substantial increase of 32.47%. This is expected to capture small and faint smoke and fires in real-time during the early stages of forest fires, even in nighttime environments. It is expected to play an important role in the development of forest fire prevention and response systems in the future, and ultimately contribute to minimizing damage caused by forest fires.

References

Al-Dabbagh, A. M., & Ilyas, M. (2023). Uni-temporal Sentinel-2 imagery for wildfire detection using deep learning semantic segmentation models. *Geomatics, Natural Hazards and Risk, 14*(1), 2196370.

Dong, H., Yuan, M., Wang, S., Zhang, L., Bao, W., Liu, Y., & Hu, Q. (2023). PHAM-YOLO: A parallel hybrid attention mechanism network for defect detection of meter in substation. *Sensors, 23*(13), 6052.

Fu, J., Zheng, H., & Mei, T. (2017). Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition. *In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Honolulu, HI, USA, 21–26 July 2017.

Jindal, P., Gupta, H., Pachauri, N., Sharma, V., & Verma, O. P. (2021). Real-time wildfire detection via image-based deep learning algorithm. In Soft Computing: Theories and Applications: *Proceedings of SoCTA 2020, Volume 2* (pp. 539-

550). Springer Singapore.

Liu, P., Xiang, P., & Lu, D. (2023). A new multi-sensor fire detection method based on LSTM networks with environmental information fusion. *Neural Computing and Applications, 35*(36), 25275-25289.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems, 32*.

Redmon, J. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Ripple, W. J., Wolf, C., Gregg, J. W., Rockström, J., Mann, M. E., Oreskes, N., ... & Crowther, T. W. (2024). The 2024 state of the climate report: Perilous times on planet Earth. *BioScience, biae087*.

Scott, J. H., Thompson, M. P., & Calkin, D. E. (2013). A wildfire risk assessment framework for land and resource management.

Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7464-7475).

Wang, C. Y., Liao, H. Y. M., & Yeh, I. H. (2022). Designing network design strategies through gradient path analysis. arXiv preprint arXiv:2211.04800.

Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., ... & Tang, X. (2017). Residual attention network for image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3156-3164).

Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. *In Proceedings of the European conference on computer vision* (ECCV) (pp. 3-19).

Yu, L., Wang, N., & Meng, X. (2005, September). Real-time forest fire detection with wireless sensor networks. *In Proceedings. 2005 International Conference on Wireless Communications, Networking and Mobile Computing, 2005*. (Vol. 2, pp. 1214-1217). IEEE.

Zope, V., Dadlani, T., Matai, A., Tembhurnikar, P., & Kalani, R. (2020, May). IoT sensor and deep neural network based wildfire prediction system. *In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 205-208). IEEE.