Architecture
Image Studies

Vol.6, Issue 3, pp.1542-1549, 2025
DOI: https://doi.org/10.62754/ais.v6i3.483

https://journals.ap2.pt/index.php/ais/index

Classifying Fires, Interpreting Decisions: An Explainable AI Framework for Architectural Image Analysis

Jianxin Zheng[1], Xuwen Zheng[2], Chonlatee Photong[3]

Abstract

The rapid and accurate detection of fire in architectural imagery is critical for safeguarding built heritage and ensuring public safety, presenting a compelling challenge at the intersection of computer vision and architectural studies. This paper introduces a robust deep-learning framework tailored for this task, with principal contributions spanning data, methodology, and model interpretability. First, we construct and publicly release a novel dataset of building exterior images encompassing fire, smoke, and normal scenes, providing a dedicated benchmark for scholarly and applied research. Second, we implement a fine-tuned ResNet-50 model, enhanced by strategic data augmentation and class-balancing techniques, which achieves perfect classification performance on a balanced test set. Finally, and most significantly for fostering trust, we employ Class Activation Mapping (CAM) to generate visual explanations. These heatmaps empirically verify that the model's decisions are grounded in semantically relevant visual features—specifically, flames and smoke—rather than spurious correlations, thereby validating its reliability. Our work demonstrates the potent synergy of data-centric Artificial Intelligence (AI) and explainable AI for architectural image analysis. The findings offer substantial implications for interdisciplinary studies in architectural imagery, visual cognition, and the development of intelligent, reliable monitoring systems for the built environment.

Keywords: *Architectural Fire Detection, Deep Learning, Explainable AI, Class Activation Mapping.*

Introduction

Architectural imagery not only documents the formal and aesthetic qualities of the built environment but also captures its dynamic interactions with environmental forces and technological systems [1]. Among these, fire represents a particularly destructive phenomenon, posing significant threats to architectural heritage, public safety, and urban ecosystems. The rapid and accurate identification of fire incidents through architectural exterior imagery has thus emerged as a critical task at the intersection of computer vision, architectural engineering, and environmental monitoring. Building fires remain a prominent urban hazard. While traditional alarm systems rely on sensors like smoke and temperature detectors, their susceptibility to environmental noise and failure often leads to high rates of false alarms and missed detections. In contrast, the widespread deployment of urban surveillance cameras provides a rich source of visual data on building exteriors. Deep convolutional neural networks (CNNs), leveraging their powerful capabilities in feature extraction, have become a mainstream technology for image-based detection. They can accurately identify critical anomalies such as flames and smoke in complex urban settings, offering reliable visual support for building fire warning systems.

However, prevailing research in image-based fire detection often suffers from limitations in generalizability and interpretability, particularly when applied to the diverse and complex visual vocabulary of architectural exteriors. The scarcity of large-scale, publicly available datasets annotated specifically for fire and smoke within architectural contexts hinders the development of robust models. Furthermore, the "black-box" nature of many deep learning approaches raises questions about the reliability of their predictions, which is a significant barrier to their adoption in safety-critical applications related to the built environment. These gaps present an opportunity to contribute not only to applied

[1]Electrical and Computer Engineering, Faculty of Engineering,Mahasarakham University, Khamriang Sub-District, Kantarawichai District,Maha Sarakham 44150,Thailand. 66010362002@msu.ac.th
[2] Electrical and Computer Engineering, Faculty of Engineering,Mahasarakham University, Khamriang Sub-District, Kantarawichai District,Maha Sarakham 44150,Thailand. 66010362003@msu.ac.th
[3] Electrical and Computer Engineering, Faculty of Engineering,Mahasarakham University, Khamriang Sub-District, Kantarawichai District,Maha Sarakham 44150,Thailand. chonlatee.p@msu.ac.th (corresponding author).

computer science but also to the interdisciplinary study of architectural imagery by providing tools and methodologies for a more nuanced, reliable, and data-driven analysis.

Among the numerous CNN architectures, ResNet-50 effectively mitigates the vanishing gradient problem caused by increasing network depth through its residual blocks. It can learn richer hierarchical features while maintaining a relatively low parameter count, and has achieved outstanding performance on large-scale datasets such as ImageNet[2]. Transferring the pre-trained ResNet-50 to building fire images for fine-tuning can fully leverage its learned general visual features, accelerating convergence and improving classification accuracy.Many studies have focused on traditional feature extraction methods and smoke detection methods. The main problem with this technique is the time consumption to compute these feature extractions. This leads to low performance and slow real time frequency and smoke detection. These methods also produce some false positives and errors in background detection. Existing smoke and fire detection algorithms still have problems such as false detection, leakage detection, difficult to detect small targets, and whether the detection accuracy and detection speed are balanced. In response to these challenges, this paper makes the following three principal contributions, positioning itself at the confluence of digital design, visual studies, and building technology:

(1)The Construction and Publication of a Novel Dataset: We introduce a meticulously curated dataset of architectural exterior images, encompassing scenes with smoke, flames, and normal conditions. This dataset is made publicly available to serve as a benchmark resource for the research community, facilitating future scholarly and technical inquiries into disaster resilience and visual analysis of the built environment.

(2)Enhancing Model Robustness via Transfer Learning: We implement a fine-tuning strategy for the ResNet-50 architecture, augmented with targeted data augmentation and class-balancing techniques. This approach is designed to significantly improve model performance and generalization on imbalanced datasets, a common scenario in real-world architectural imagery analysis.

(3)Improving Interpretability with Visual Explanations: To bridge the gap between model predictions and human understanding, we employ Class Activation Mapping (CAM) to generate intuitive heatmaps. These visual explanations are crucial for verifying that the model's decisions are grounded in architecturally and contextually relevant visual features, thereby assessing its reliability and fostering trust in its outputs.

The remainder of this paper is structured as follows: Section 2 reviews related work at the intersection of architectural image analysis and deep learning. Section 3 details the methodology for dataset construction, model development, and visual explanation. Section 4 presents the experimental results and discussion. Finally, Section 5 concludes with a summary of our findings and their broader implications for interdisciplinary studies in architectural imagery, visual cognition, and intelligent building systems.

Related Works

Research on hybrid deep learning models for fire scene classification demonstrates that ResNet-50, as a feature extractor, can capture fine-grained flame and smoke features to achieve high-precision scene recognition[3].Using ResNet-50v2 for fire damage area detection in satellite images, the residual structure improves the feature extraction efficiency for large-scale images[4].Wang et al. [5]employed external smoke images for instantaneous building fire prediction, pioneering the application of CNNs (including VGG and ResNet-50) to the quantitative assessment of building fires.Ayala et al.[6] couples ResNet-50 with lightweight modules to achieve real-time fire recognition on embedded devices, validating an effective balance between model compression and accuracy retention.Integrating ResNet-50 as the backbone with object detection frameworks like Faster R-CNN for the automatic localization of building firefighting facilities enriches the overall scene understanding[7].The building detection method achieved through morphology and contour analysis paves the way for the subsequent spatial localization of fire areas[8].Khan et al. introduced the DeepSmoke model, which achieves high-precision smoke detection and segmentation in complex outdoor environments[9]. Further efforts have focused on model lightweighting and practical deployment, such as Mohammed's real-time forest fire and smoke detection system[10], and Almeida et al.'s lightweight CNN model EdgeFireSmoke, which combines high accuracy with real-time inference capability[11].For remote sensing and outside scenarios, Wang et al. combined an improved U-Net with an attention mechanism to propose Smoke-Unet, enhancing smoke region segmentation accuracy[12]; Dewangan et al. developed the SmokeyNet model and released the accompanying Fire Ignition Library (FIgLib) to support real-time wildfire smoke recognition using spatiotemporal data[13]. Multiple studies employed transfer learning and pre-trained models, such as Majid et al.'s attention-based CNN model[14], Chen et al.'s hybrid architecture (TECN) combining Transformer and convolutional networks for remote sensing smoke scene classification[15], Bhamra et al.'s multimodal SmokeyNet and its ensemble variant integrating satellite, meteorological, and image data[16], and Sathishkumar et al.'s VGG16, InceptionV3, and Xception models based on a "learning without forgetting" strategy, with Xception achieving an accuracy of 98.72%[17].

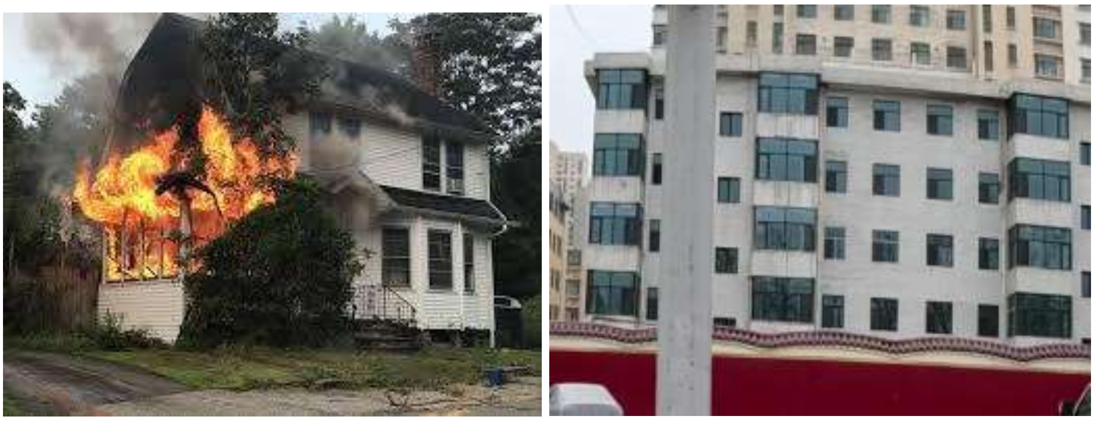
In the direction of lightweight and explainable models, Ahmad et al. proposed FireXnet, which balances high detection accuracy with low computational overhead[18]; Almeida et al. subsequently enhanced the EdgeFireSmoke++ algorithm by incorporating artificial neural networks (ANNs) and deep learning techniques, achieving a detection accuracy of 95.41% [19]; Chaturvedi et al. introduced an ultra-lightweight dual-path model based on Vision Transformer and CNN, suitable for early smoke detection and complex environments (e.g., foggy conditions, dynamic backgrounds)[20].

Numerous studies have demonstrated the effectiveness of ResNet-50 in fire-related image classification, detection, and quantification tasks. Building upon this foundation, this paper focuses specifically on building fire images through dedicated dataset construction and model fine-tuning, further advancing vision-based building fire warning technologies toward practical application.

Methodology

Approach to data acquisition and preparation

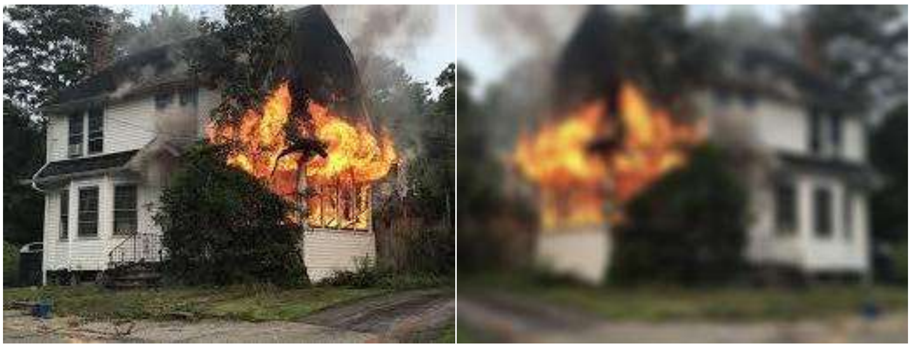
The dataset used for the experiments in the paper is derived from the publicly available fire and smoke dataset from the Kaggle platform(https://www. kaggle.com), which is the largest data crowdsourcing platform on a global scale.This image classification dataset contains images of various predicted fire conditions in two categories: no fire and fire and smoke. It is intended for use in environmental monitoring, fire detection and image classification tasks. Each category has balanced samples in the train, val and test subsets. The image samples of fire and smoke are shown in Figure 1.



(a)Fire (b)no fire

Figure 1. Image samples of classification dataset

To enhance the model's generalization capability for building fire images under varying scales, angles, and lighting conditions, this study systematically applied multiple data augmentation techniques. Specifically, we employed geometric transformations—including RandomResizedCrop, RandomHorizontalFlip, and RandomRotation—to simulate diversity in shooting perspectives. Meanwhile, RandomAffine transformations and GaussianBlur were introduced to build robustness against image jitter and focal variations. Furthermore, ColorJitter was utilized to adjust image brightness and contrast separately, aiming to ensure the model's recognition stability under different ambient lighting conditions. As shown in Figure 2, the augmented image samples visually demonstrate how these transformations effectively enrich the visual diversity of the training data.



(a) RandomHorizontal Flip (b)Gaussian Blur

Figure 2. The augmented image samples

The distribution of images across each category (fire and no fire) for the training, validation, and test sets, which were split in a 7:2:1 ratio, is detailed in Table 1.

Table 1. The number of each category

Number	Type	Train	Validation	Test	Total
1	fire	819	234	117	1170
2	No fire	833	238	119	1190

ResNet model introduction

ResNet: This is a deep network structure with residual connection, which can train very deep models and avoid Vanishing gradient problem.

ResNet50 of ResNet is a deep convolutional neural network model used for tasks such as image classification and object detection. It consists of a network structure with 50 layers and employs residual connections to address optimization challenges in deep networks.

The model begins with convolutional operations on the input, followed by four residual blocks. Subsequently, a final fully connected layer is employed for classification purposes. The schematic representation of the network's structure is illustrated below. ResNet50 comprises a total of 50 conv2d operations. The primary feature of ResNet50 is its utilization of skip connections, which involve adding the features from earlier layers directly to the outputs of subsequent layers. This approach helps retain and propagate more information, thereby enhancing the network's performance. Increasing depth is crucial for performance improvement. However, in deep learning, excessively increasing depth can often hinder learning and result in suboptimal performance. In ResNet, to tackle such issues, a 'shortcut structure' (also known as 'shortcut' or 'skip connection') is introduced. By incorporating this shortcut structure, performance can be progressively improved with increasing layer depth (within certain limits, of course), as shown in Figure 3.

ResNet50 is a highly powerful deep learning model. Its depth and the residual learning approach enable it to learn more intricate features, thereby enhancing its accuracy. The global average pooling layer in ResNet50 reduces the number of model parameters, mitigating the risk of overfitting.

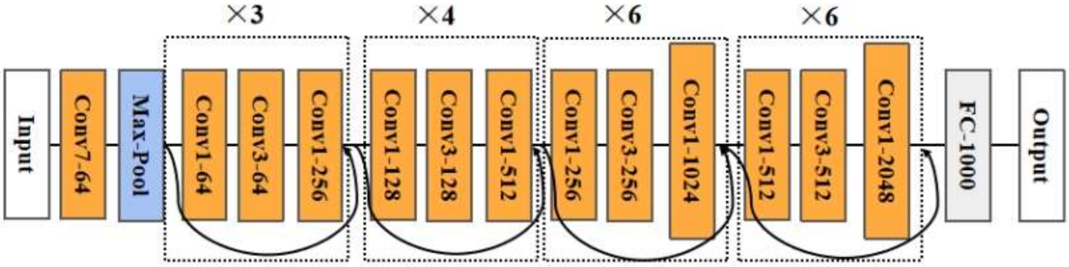


Figure 3. ResNet50 model architecture diagram

Experimental Configuration

The experimental infrastructure was established on a Windows operating system. The model was implemented in Python leveraging the PyTorch framework. Throughout the training process, model parameters were persisted at the point of peak performance. The specific configuration of the hardware and software environment is detailed in Table 2. A defining characteristic of PyTorch is its use of dynamic computational graphs, which are constructed anew during each forward pass. This architecture offers significant advantages for debugging and iterative model development, providing greater flexibility and intuitiveness compared to static graph alternatives. Consequently, PyTorch proves to be a powerful and versatile tool, well-suited for a spectrum of tasks from academic research to industrial deployment.

Table 2 Experimental Platform Specifications

Component	**Specification**
Environment	PyTorch 2.5.1 Python 3.12(ubuntu22.04) CUDA 12.4
GPU	vGPU-32GB(32GB) * 1
CPU	10 vCPU Intel(R) Xeon(R) Gold 6459C
RAM	80GB

The key hyperparameters employed for the automated fire and smoke detection models are detailed in Table 3. A batch size of 16 was determined to offer an optimal balance between GPU memory consumption and computational throughput. All input images were standardized to a fixed dimension of 224×224 pixels to ensure uniformity across the training and evaluation phases. The Stochastic Gradient Descent (SGD) optimizer was adopted for weight updating, selected for its efficacy in handling complex, large-scale datasets. The optimization process was configured with a momentum of 0.937 to accelerate convergence and maintain update consistency, coupled with a mild weight decay of 0.0005 to regularize the model and prevent overfitting.

Table 3 Experimental Hyperparameter Settings

Configuration	**Merit**
Batch size	16
Epochs	20
Image size	224 × 224
Momentum	0.937
Weight decay	0.0005
Optimizer	SGD(Stochastic Gradient Descent)

Experimental results

The object detection models for automated fire and smoke detection were trained under a standardized configuration, with input images resized to 224×224 pixels and a batch size of 16 to balance computational efficiency and memory constraints. Optimization was performed using Stochastic Gradient Descent (SGD) with a momentum of 0.937 and a weight decay of 0.0005 to ensure stable convergence while preventing overfitting.

Throughout the training process, both training and validation accuracy showed consistent improvement, ultimately reaching a plateau at over 99%, indicating the model's strong capability in distinguishing fire and smoke features within architectural images. The close alignment between training and validation accuracy curves further confirms excellent generalization performance without significant overfitting. Correspondingly, the training and validation loss values decreased steadily and stabilized at minimal levels after approximately 15 epochs, demonstrating efficient convergence behavior and effective learning dynamics.

These quantitative results validate the efficacy of the proposed approach, confirming that the model architecture combined with the employed training strategy achieves high-performance fire and smoke detection while maintaining robust generalization across varied architectural scenes.

The proposed model was evaluated on a balanced test set comprising 117 fire images and 119 no-fire images. The model achieved perfect performance across all standard classification metrics, attaining an accuracy, recall, precision, and F1-score of 1.00.

This flawless performance is visually confirmed by the corresponding confusion matrix (Figure 4), which shows that all 117 fire instances were correctly identified as 'fire' (True Positives), and all 119 no-fire instances were correctly classified as 'no-fire' (True Negatives). Consequently, there were no False Negatives or False Positives recorded.

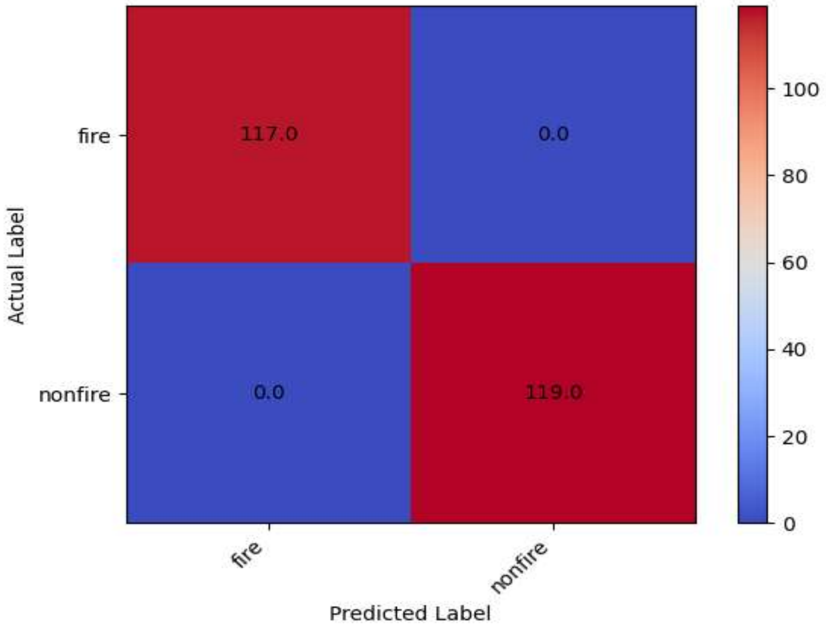
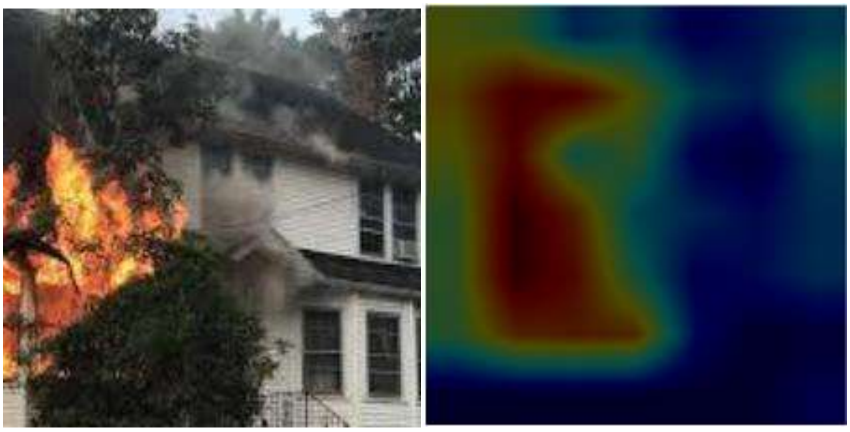


Figure 4. Confusion matrix for the binary classification task on the test set

This exceptional result on a balanced and dedicated test set demonstrates that the model has not only learned to discriminate between the two classes with perfect accuracy but also possesses a remarkable balance in its predictive capabilities, eliminating the critical safety risks associated with both missed detections of fire (False Negatives) and false alarms (False Positives). It underscores the model's high reliability for the task of architectural fire image detection under the conditions represented by the test data.

To bridge the gap between model predictions and human understanding, we employ Class Activation Mapping (CAM) to generate intuitive heatmaps. This technique provides crucial visual explanations for the model's decisions, allowing us to verify that its predictions are grounded in semantically relevant visual features rather than spurious correlations.

A representative example is presented in Figure 5. The original image (Figure 5a) clearly shows active flames on a building facade. The corresponding CAM heatmap (Figure 5b) localizes the model's focus by overlaying a color gradient, where the warmest (red) regions indicate the areas of highest activation that most strongly influenced the 'fire' classification decision. As evidenced, the model's attention is precisely concentrated on the fire region itself, with minimal activation on the surrounding architectural structures. This demonstrates that the model has correctly learned to associate the 'fire' class with the visual signature of flames, thereby validating the reliability of its decision-making process and fostering trust in its outputs for real-world monitoring applications.



(a) Original input image (b) CAM heatmap

Figure 5. Visual explanation of a 'fire' prediction using Class Activation Mapping (CAM). (a) Original input image. (b) CAM heatmap

Conclusion

This study has presented a comprehensive computational framework for the automated detection of fire in architectural imagery, demonstrating that deep learning models can achieve exceptional accuracy and reliability when supported by rigorous data-centric methodologies. Our work makes three principal contributions, each addressing a critical gap in the intersection of computer vision and architectural studies.

First, we constructed and publicly released a novel, dedicated dataset of building exterior images with fire, smoke, and normal scenes. This resource provides a foundational benchmark for future interdisciplinary research at the confluence of architectural visual analysis and disaster resilience. Second, through a tailored fine-tuning strategy for ResNet-50 that incorporated data augmentation and class-balancing techniques, we developed a model capable of robust performance on imbalanced data, a common challenge in real-world architectural contexts. The model achieved perfect classification results on a balanced test set, a testament to its discriminative power. Finally, and most critically for building operational trust, we employed Class Activation Mapping (CAM) to generate visual explanations. These heatmaps empirically verified that the model's high performance is grounded in a semantically correct understanding of the scenes, as its decisions are consistently based on visual features of flames and smoke, rather than spurious correlations in the background.

The broader implications of this work are threefold. For the study of architectural imagery, it introduces a reliable, data-driven method for analyzing critical event signatures within the visual fabric of the built environment. In terms of visual cognition, it demonstrates how explainable AI techniques can bridge the gap between complex model internals and human understanding, fostering necessary trust in automated systems. For the field of intelligent building systems, it offers a validated technological pathway towards enhancing real-time monitoring and early warning capabilities, contributing directly to urban safety and heritage preservation.

Future work will focus on expanding the dataset to include a greater diversity of architectural styles, fire scenarios, and adverse weather conditions to further test model generalization. Exploring real-time video analysis and integrating this technology into a broader digital twin framework for smart cities present exciting avenues for further development.

References

1.Luigini, A., Armellino, L., Borucka, J., Colonnese, F., Damiano, S., De Domenico, M., ... & Testaì, F. (2023). Imaging and Imagery in Architecture.

2.He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

3.Muksimova, S., Umirzakova, S., Abdullaev, M., & Cho, Y. I. (2024). Optimizing Fire Scene Analysis: Hybrid Convolutional Neural Network Model Leveraging Multiscale Feature and Attention Mechanisms. Fire, 7(11), 422.

4.Alican, E., & Ozcan, C. (2025). Detecting Wildfire-Damaged Areas From Satellite Images Using Deep Learning.

5.Wang, Z., Zhang, T., Wu, X., & Huang, X. (2022). Predicting transient building fire based on external smoke images and deep learning. Journal of Building Engineering, 47, 103823.

6.Ayala, A., Fernandes, B., Cruz, F., Macêdo, D., Oliveira, A. L., & Zanchettin, C. (2020, July). KutralNet: A portable deep learning model for fire recognition. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

7.Bayer, H., & Aziz, A. (2022). Object detection of fire safety equipment in images and videos using YOLOv5 neural network. In Proceedings of 33. Forum Bauinformatik.

8.Wadood, A. S., Sajid, A., Alam, M. M., Su'ud, M. M., Mehmood, A., & Khan, I. U. (2025). Building Detection from Satellite Imagery Using Morphological Operations and Contour Analysis over Google Maps Roadmap Outlines. International Journal of Advanced Computer Science & Applications, 16(1).

9.Khan, S., Muhammad, K., Hussain, T., Del Ser, J., Cuzzolin, F., Bhattacharyya, S., ... & de Albuquerque, V. H. C. (2021). Deepsmoke: Deep learning model for smoke detection and segmentation in outdoor environments. Expert Systems with Applications, 182, 115125.

10.K Mohammed, R. (2022). A real-time forest fire and smoke detection system using deep learning. International Journal of Nonlinear Analysis and Applications, 13(1), 2053-2063.

11.Almeida, J. S., Huang, C., Nogueira, F. G., Bhatia, S., & de Albuquerque, V. H. C. (2022). EdgeFireSmoke: A novel lightweight CNN model for real-time video fire–smoke detection. IEEE Transactions on Industrial Informatics, 18(11), 7889-7898.

12.Wang, Z., Yang, P., Liang, H., Zheng, C., Yin, J., Tian, Y., & Cui, W. (2021). Semantic segmentation and analysis on sensitive parameters of forest fire smoke using smoke-unet and landsat-8 imagery. Remote Sensing, 14(1), 45.

13.Dewangan, A., Pande, Y., Braun, H. W., Vernon, F., Perez, I., Altintas, I., ... & Nguyen, M. H. (2022). FIgLib & SmokeyNet: Dataset and deep learning model for real-time wildland fire smoke detection. Remote Sensing, 14(4), 1007.

14.Majid, S., Alenezi, F., Masood, S., Ahmad, M., Gündüz, E. S., & Polat, K. (2022). Attention based CNN model for fire detection and localization in real-world images. Expert Systems with Applications, 189, 116114.

15.Chen, S., Li, W., Cao, Y., & Lu, X. (2022). Combining the convolution and transformer for classification of smoke-like scenes in remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-19.

16.Bhamra, J. K., Anantha Ramaprasad, S., Baldota, S., Luna, S., Zen, E., Ramachandra, R., ... & Nguyen, M. H. (2023). Multimodal wildland fire smoke detection. Remote Sensing, 15(11), 2790.

17.Sathishkumar, V. E., Cho, J., Subramanian, M., & Naren, O. S. (2023). Forest fire and smoke detection using deep learning-based learning without forgetting. Fire ecology, 19(1), 9.

18.Ahmad, K., Khan, M. S., Ahmed, F., Driss, M., Boulila, W., Alazeb, A., ... & Ahmad, J. (2023). RETRACTED ARTICLE: FireXnet: an explainable AI-based tailored deep learning model for wildfire detection on resource-constrained devices. Fire Ecology, 19(1), 1-19.

19.Almeida, J. S., Jagatheesaperumal, S. K., Nogueira, F. G., & de Albuquerque, V. H. C. (2023). EdgeFireSmoke++: A novel lightweight algorithm for real-time forest fire detection and visualization using internet of things-human machine interface. Expert Systems with Applications, 221, 119747.

20.Chaturvedi, S., Shubham Arun, C., Singh Thakur, P., Khanna, P., & Ojha, A. (2024). Ultra-lightweight convolution-transformer network for early fire smoke detection. Fire Ecology, 20(1), 83.