SOA Deep Learning

Forests play a crucial role in the ecosystem, providing a habitat for numerous plant and animal species. However, the escalating threat of forest fires due to changing weather patterns and rising temperatures caused by global warming poses a significant risk to biodiversity. Early detection of these fires is paramount to mitigating their impact and minimizing damage. In our research, we use raw data captured by onsite PTZ cameras to obtain high-resolution live images of the forest. We aim to use state-of-the-art models like YOLO V9 and CLIP for early detection of Forest Fire and Smoke under normal and Hazy conditions.

Numerous studies worldwide have delved into the realm of forest fire detection using computer vision techniques, encompassing segmentation, object detection, and bounding box methodologies. Among the prevalent algorithms for feature extraction in computer vision, Convolutional Neural Networks (CNNs) stand out as a cornerstone. For instance, [1] employs an ensemble of deep CNNs, hierarchically extracting features from the data. Meanwhile, [2] adopts the MobileNetV3 architecture within CNNs for fire detection, boasting a commendable 90% accuracy rate.

In another notable advancement, Islam et al. [3] leverage a pre-trained EfficientnetB7 model integrated with attention mechanisms and Bayesian optimization, achieving an impressive 97% accuracy and 98% F1 score in forest fire detection.

The YOLO (You Only Look Once) architecture, renowned for its efficiency in object detection tasks, has also gained attention in the realm of forest fire and smoke detection. Kim et al. [4] pioneer the use of convolutional attention within YOLO V7, coupled with a bidirectional feature pyramid network, to detect forest fires and smoke at early stages with a remarkable 91.5% accuracy rate. Similarly, [5] proposes a Convolutional Neural Network (CNN) utilizing the CSPdarknet53 architecture derived from YOLO, featuring an SPP layer for smoke detection, achieving an outstanding F1 score of 97.9%. Since fire and smoke features are similar across different domains and are easily transferable to forest conditions hence most of the researchers finetune pre-trained models for their experiments.

For Early detection of forest fires, smoke detection plays a crucial role. However, most smoke detection algorithms have high false positive rates as they confuse clouds with smoke due to high spatial correspondence. Nikolay et al [9] devise a unique solution by separating the sky from the image by applying a Gaussian filter that Identifies the line of the horizon as the line that has the maximum value of gradients effectively reducing the false positive rates.

Moreover, Contrastive Language-Image Pretraining (CLIP) significantly advances image classification through zero-shot learning. Research conducted by [6] demonstrates CLIP's remarkable 95% accuracy in ImageNet’s Top-5 classification, achieved without direct training on the dataset. While CLIP has not yet been specifically applied to forest fire detection, its image classification capabilities offer promise for identifying such incidents within frames.

Although CLIP has traditionally been employed for image classification, studies such as [7], have explored its potential for object detection through a variant model known as GridCLIP. This variant facilitates the recognition of finer details within images by employing a grid-level representation. Notably, GridCLIP achieves training and testing speeds that are 43 and 5 times faster, respectively, compared to state-of-the-art models, while delivering comparable performance on the LVIS benchmark.

Moreover, [8] introduces a approach termed CLIP-Driven Image Segmentation, utilising a variant named CRIS. This model integrates semantic information from textual descriptions into visual representations, thereby enhancing precision in image segmentation tasks. Notably, CRIS surpasses previous benchmarks, achieving an impressive accuracy of 70.47% on the COCO benchmark.

While CLIP initially focused on image classification tasks, these aforementioned variants extend its applicability to object detection and image segmentation without necessitating additional training. This versatility holds significant potential for forest fire detection or segmentation applications, leveraging CLIP's inherent zero-shot capabilities and obviating the need for extensive local dataset training.

[1] Yu, Y.; Yao, M. When Convolutional Neural Networks Meet Laser-Induced Breakdown Spectroscopy: End-to-End Quantitative Analysis Modeling of ChemCam Spectral Data for Major Elements Based on Ensemble Convolutional Neural Networks. *Remote Sens.* **2023**

[2] Li, Y.; Zhang, W.; Liu, Y.; Jin, Y. A Visualized Fire Detection Method Based on Convolutional Neural Network beyond Anchor. *Appl.*

[3] Islam, A.M.; Masud, F.B.; Ahmed, M.R.; Jafar, A.I.; Ullah, J.R.; Islam, S.; Shatabda, S.; Islam, A.K.M.M. An Attention-Guided Deep-Learning-Based Network with Bayesian Optimization for Forest Fire Classification and Localization. *Forests* **2023**,

[4]. Kim, S.-Y.; Muminov, A. Forest Fire Smoke Detection Based on Deep Learning Approaches and Unmanned Aerial Vehicle Images. *Sensors* **2023**

[5] Huo, Y.; Zhang, Q.; Jia, Y.; Liu, D.; Guan, J.; Lin, G.; Zhang, Y. A Deep Separable Convolutional Neural Network for Multiscale Image-Based Smoke Detection. *Fire Technol.* **2022**

[6] Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sutskever, I. Learning Transferable Visual Models From Natural Language Supervision. *arXiv preprint arXiv:2103.00020.* **2021**

[7] Lin, J.; Gong, S. GridCLIP: One-Stage Object Detection by Grid-Level CLIP Representation Learning. *arXiv preprint arXiv:2303.09252v1.* **2023**

[8] Wang, Z.; Lu, Y.; Li, Q.; Tao, X.; Guo, Y.; Gong, M.; Liu, T. CRIS: CLIP-Driven Referring Image Segmentation. *arXiv preprint arXiv:2110.12009.* **2021**

[9] Abramov, Nikolay et al. Intelligent Methods for Forest Fire Detection Using Unmanned Aerial Vehicles. *https://www.mdpi.com/2571-6255/7/3/89#B14-fire-07-0008.* **2024**