Improving Deep Learning-Based Wildfire Smoke Plume Detection with a Multi-Model Ensemble Approach

Author Name
Institution
email@example.com

Abstract

With the increasing frequency and severity of wildfires, there is an urgent need for effective and rapid wildfire and smoke detection tools. Recent advancements in computer vision have demonstrated the potential of deep learning models, particularly neural networks, to automate the partitioning of high-resolution images into labelled segments. However, single-model approaches can struggle with generalization and accuracy in diverse conditions. To address these challenges, we propose creating an ensemble of deep learning models to produce accurate and representative predictions of wildfire smoke plumes and their relative density (light, medium, heavy) in Geostationary Operational Environmental Satellite (GOES) imagery. Our results indicate that this ensemble technique can significantly improve performance compared to using a single model. This multi-model ensemble is expected to support fire and hazard management by being able to automate the monitoring of smoke in real-time from satellite imagery, providing a valuable tool for air quality and fire hazard management in the face of worsening wildfires.

1 Introduction

Wildfires increase smoke and particulate matter in the atmosphere, posing more risks of respiratory issues and other air quality-induced health problems in recent years (Burke et al., 2021). Effective and timely wildfire and smoke detection tools are thus essential for supporting hazard management and mitigating risks to human health.

The National Oceanic and Atmospheric Administration (NOAA) Geostationary Operational Environmental Satellites (GOES) provide high spatial and temporal resolution imagery of North America (Goodman et al., 2019), which can be leveraged to detect the presence and density of smoke plumes. The Hazard Mapping System (HMS) Fire and Smoke Product currently relies on human analysts to annotate the presence of smoke over North America using imagery from the GOES imagery (McNamara et al., 2004). However, this product is limited by the availability of human analysts, outputting annotations once to several times a day, and usually has a delay between smoke occurance and the annotation. To address this limitation in the availability of smoke data, we are leveraging advancements in deep learning to automate the detection of smoke from GOES imagery. Deep learning models, particularly encoder-decoder neural networks, have shown promise in automating the semantic segmentation (labelling images on a pixel-wise basis with multiple classes) of high-resolution images (Minaee et al., 2022). By automating this task, we can enable more frequent and consistent detection of smoke plumes.

This proposal focuses on enhancing the capability of smoke detection with deep learning through the use of multi-model ensemble techniques. It has been shown for classification that ensemble methods that combine the predictions of multiple classifers can often perform better than a single classifer (Dietterich, 2000). Particularly, utilizing a diverse set of classifiers in an ensemble is important to achieve the improvement in performance (Kuncheva and Whitaker, 2003). In the neural network setting, combining the predictions of multiple independently-trained models can improve generalization and prediction accuracy of neural networks (Hansen and Salamon, 1990), (Cheng Ju and van der Laan, 2018), (Giacinto and Roli, 2001). This approach aims to provide a more reliable and accurate tool for real-time monitoring of smoke, ultimately supporting fire and hazard management efforts and contributing to climate resilience and adaptation strategies.

2 Data and Methods

The dataset we use consists of 183,672 samples, each with three spectral channels (C01-C03) of GOES imagery paired with HMS smoke annotations (pixel-wise labels of smoke density of light, medium, or heavy) for a specific datetime and location. The data spans 2018-2024, and we set aside 2023 for validation and 2022 for testing, with the remaining years used for training.

We are utilizing a variety of pre-developed encoder-decoder architectures that were designed for semantic segmentation contained within the Segmentation Models package (Iakubovskii, 2019). These architectures include different features such as multi-scale fields-of-view and precise boundary localization (Chen et al., 2018), (Li et al., 2018), (Zhou et al., 2018), which are important for accurately detecting smoke plumes that can vary in size and appearance. A collection of these models are trained independently on 8 Nvidia P100 GPUs using the Adam optimizer with a learning rate of 0.001 and batch size of 128. The ensemble method we are using is an unweighted average of the predictions of the models. This method is straightforward to implement and has been shown to be effective in practice (Cheng Ju and van der Laan, 2018). This ensemble framework is shown in Figure 1.

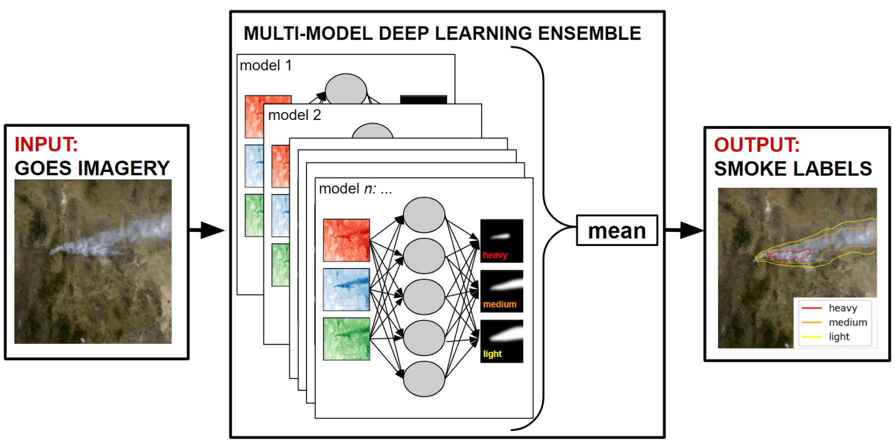

Figure 1: Multi-Model Ensemble Framework.

3 Results

We measure model performance in terms of Intersection over Union (IoU) score (Equation 1) which quantifies the alignment between the model predicted and the ground truth.

$$\mathrm{IoU}_{\text{overall}} = \frac{\sum_{i=\text{light}}^{\text{heavy}} |y_i \cap y_i^*|}{\sum_{i=\text{light}}^{\text{heavy}} |y_i| \cup |y_i^*|} \tag{1}$$

- add a sample demonstrating the smoothing of boundaries in ensemble compared to individual model preds

Architecture	Heavy IoU	Medium IoU	Light IoU	Overall IoU
DeepLabV3Plus	0.347077	0.441486	0.666248	0.59859
PSPNet	0.374195	0.482493	0.651067	0.59614
LinkNet	0.360252	0.469944	0.620847	0.570342
Ensemble (N=8)	0.399515	0.507199	0.691542	0.634591

Table 1: IoU results across three classes of smoke (light, medium, heavy) and over all densities. Prsented for different individual models of different architectures, along with the ensemble (N=8) performance.

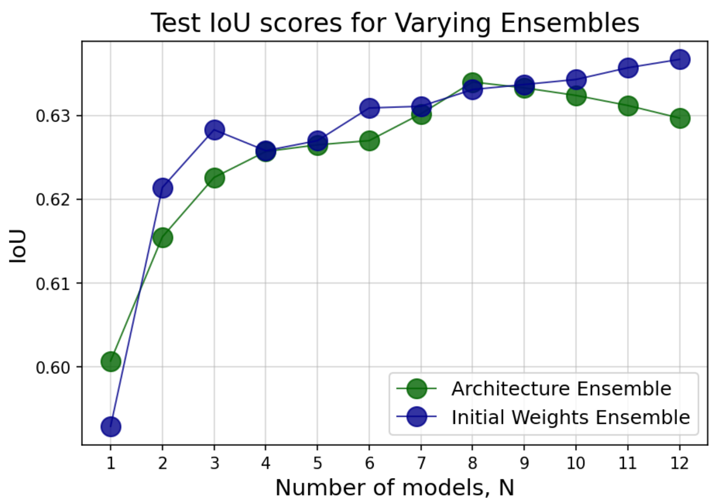

Figure 2: Ensemble IoU Performance over all smoke densities as a function of the number of models in the ensemble. Three different schemes of ensemble design are shown.

- discuss: made combinations of ensembles in terms of best individual score

- discuss: how the ensemble improves performance over the best individual model

- why is there a plateau in the current ensemble size plot? How can I explore potential explanations for this?

4 Conclusions and Future Work

- We created an ensemble of architecturally-diverse deep learning models that improves on Test IoU and smoothes jagged boundaries.

- This multi-model data-driven ensemble can be used to automate the monitoring of smoke from GOES imagery.

- We are experimenting with more ensemble techniques (e.g. regional modelsz) to improve performance.

- climate change implications of these results...

References

Marshall Burke, Anne Driscoll, Sam Heft-Neal, Jiani Xue, Jennifer Burney, and Michael Wara. The changing risk and burden of wildfire in the united states. *Proceedings of the National Academy of Sciences*, 118(2):e2011048118, 2021. doi: 10.1073/pnas.2011048118. URL https://www.pnas.org/doi/abs/10.1073/pnas.2011048118.

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation, 2018. URL https://arxiv.org/abs/1802.02611.

Aurélien Bibaut Cheng Ju and Mark van der Laan. The relative performance of ensemble methods with deep convolutional neural networks for image classification. *Journal of Applied Statistics*, 45(15):2800–2818, 2018. doi: 10.1080/02664763.2018.1441383. URL https://doi.org/10.1080/02664763.2018.1441383. PMID: 31631918.

Thomas G. Dietterich. Ensemble methods in machine learning. *Multiple Classifier Systems*, pages 1–15, 2000.

Giorgio Giacinto and Fabio Roli. Design of effective neural network ensembles for image classification purposes. *Image and Vision Computing*, 19(9):699–707, 2001. ISSN 0262-8856. doi: https://doi.org/10.1016/S0262-8856(01)00045-2. URL https://www.sciencedirect.com/science/article/pii/S0262885601000452.

S. J. Goodman, T. J. Schmit, J. Daniels, and R. J. Redmon. *The GOES-R Series: A New Generation of Geostationary Environmental Satellites.* Elsevier, 2019.

L.K. Hansen and P. Salamon. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10):993–1001, 1990. doi: 10.1109/34.58871.

Pavel Iakubovskii. Segmentation models pytorch. https://github.com/qubvel/segmentation_models.pytorch, 2019.

Ludmila I. Kuncheva and Christopher J. Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, 51 (2):181–207, 2003. ISSN 1573-0565. doi: 10.1023/A:1022859003006. URL https://doi.org/10.1023/A:1022859003006.

Hanchao Li, Pengfei Xiong, Jie An, and Lingxue Wang. Pyramid attention network for semantic segmentation. *CoRR*, abs/1805.10180, 2018. URL http://arxiv.org/abs/1805.10180.

Donna McNamara, George Stephens, Mark Ruminski, and Tim Kasheta. The hazard mapping system (hms) - noaa's multi-sensor fire and smoke detection program using environmental satellites. *Conference on Satellite Meteorology and Oceanography*, 01 2004.

Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3523–3542, 2022. doi: 10.1109/TPAMI.2021.3059968.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. *CoRR*, abs/1807.10165, 2018. URL http://arxiv.org/abs/1807.10165.