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. Additionally, we selected one of these architectures and trained it with 12 different random seeds to generate different initial random weights before training. These models are trained independently for 24 hours on 8 Nvidia P100 GPUs using the Adam optimizer, a learning rate of 0.001, a binary cross entropy loss function, and batch size of 128. Each model gets selected based on its best validation IoU score.

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. To preliminary test what combinations of models would improve on testing performance, we experimented with different ensemble sizes (1-12 models) for combinations of different architectures and with different initial seeds.

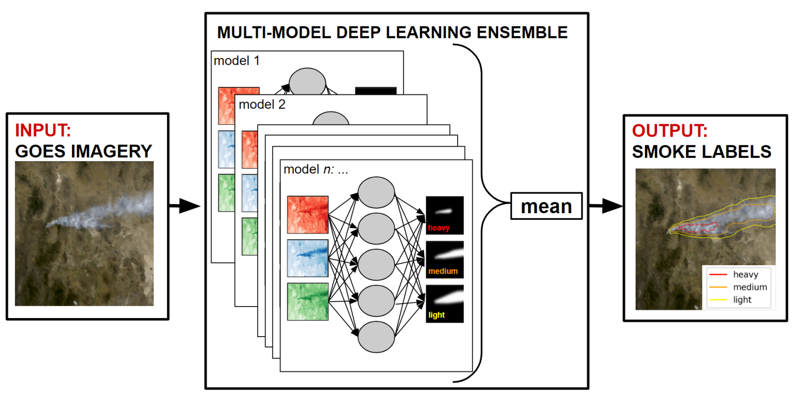

Figure 1: Multi-Model Ensemble Framework.

3 Results

We measure model performance in terms of Intersection over Union (IoU) score (Equation 1) which quantifies the alignment between the model prediction (y_i^*) and the ground truth (y_i). Table 1 shows the IoU scores for individual models and ensembles. The ensemble of 8 models with different architectures outperforms the individual models, with an improvement in the IoU score for all densities of smoke and over all densities. The ensemble of 8 models with different initial weights also outperforms the individual models, with a similar improvement in the IoU score. Figure 2 shows the IoU performance over all smoke densities as a function of ensemble size for the two ensemble schemes. The ensemble of 8 models with different initial weights improves with more models in the ensemble. This improvement is likely due to the different initializations allowing the models to search different parts of the parameter space and find different minima of the loss. The ensemble of 8 models with different architectures improves with more models in the ensemble up to 8 models, but then plateaus in performance. This plateau in performance could be due to the additional architectures not being as well suited for the task, or the models not having enough diversity in their errors to improve ensemble performance.

Figure 3 shows an example of smoke plume detection from the testing dataset. The ensemble predictions show smoother boundaries, making the prediction more comparable to the human-drawn polygon annotations.

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

	Heavy	Medium	Light	Overall
Single Model: DLV3P	0.347	0.441	0.666	0.599
Single Model: PAN	0.349	0.478	0.664	0.604
Architecture Ensemble (N=8)	0.400	0.507	0.692	0.635
Random Initial Weights Ensemble (N=8)	0.409	0.512	0.684	0.631

Table 1: IoU results across three classes of smoke (light, medium, heavy) and over all densities. Presented for different individual models of different architectures ((Chen et al., 2018); (Li et al., 2018)), along with the archiecture-based ensemble and random initial weights ensemble performance, where N denotes the number of models in the ensemble.

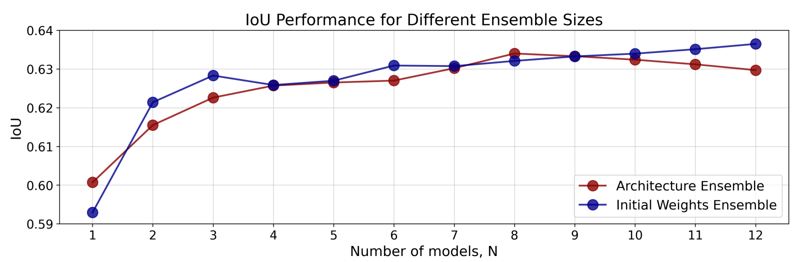

Figure 2: Ensemble IoU Performance over all smoke densities as a function of ensemble size. Presented for two different schemes of ensemble design: random initial weights (blue) and architecure-based (red).

4 Conclusions and Future Work

We explored two schemes for making ensembles of deep learning models that both improve on testing set IoU and make the predictions more realistic. There is more to be investigates, such as why the architecture-based ensemble plateaus in performance after 8 models, and how a combination of the two ensemble schemes will perform. We are also experimenting with other ensemble techniques, such as regionally-trained models, to improve detection of smoke. The ensemble of deep learning models is expected to support fire and hazard management by automating the monitoring of smoke in real-time from satellite imagery. This tool can be used to provide more frequent and consistent detection of smoke plumes, ultimately supporting climate resilience and adaptation strategies.

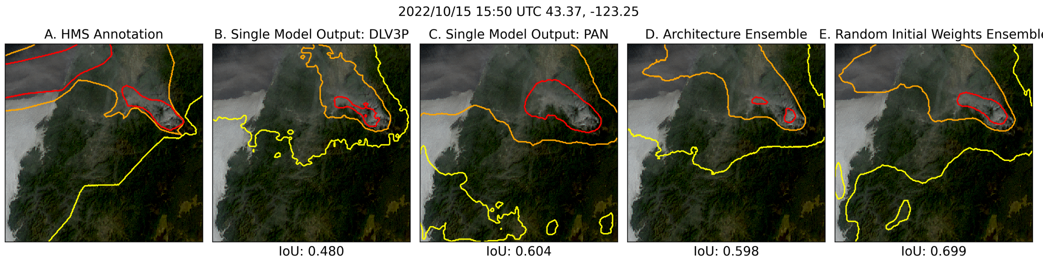

Figure 3: Example of smoke plume detection from GOES imagery with two individual models and two ensembles. Panel A displays the ground truth HMS annotation; Panels B-C show the predictions of two individual models; Panel D shows the prediction of an architecture-based ensemble of 8 models; Panel E shows the prediction of an ensemble made with 8 models with different random initial weights.

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`.