Robust Multi-Campaign Imaging Spectrometer Methane Plume Detection using Deep Learning

Jake Lee[1], Brian Bue[1], Michael Garay[1], Andrew Thorpe[1], Riley Duren[2,3], Daniel Cusworth[2,3], Alana Ayasse[2,3]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA [2]University of Arizona, Tucson, AZ [3]Carbon Mapper, Inc., Pasadena, CA

Introduction

Identification of global methane sources is critical to the quantification and mitigation of this greenhouse gas. Future spaceborne imaging spectrometer missions, such as Carbon Mapper, will provide global observations that will enable accurate mapping of methane sources. Due to the sheer data volume of these missions, manual methane source identification is infeasible, and an automated source detection method is needed. Recent works have demonstrated the feasibility of Convolutional Neural Networks (CNNs) for plume detection; however, these models have suffered from high false positive rates, noisy training data, small sample sizes, and were limited in their training and evaluation to individual flight campaigns.

To develop a more robust methane source detector,

- **We curated three AVIRIS-NG campaigns for machine learning.**
- **We trained CNN methane plume detection models on individual and collective campaigns.**
- **We evaluated detector performance within and across campaigns to determine generalization performance.**
- **We converted the CNN into an FCN to generate flightline-level methane plume saliency maps.**

Data Preparation

We curated three AVIRIS-NG campaigns:

- 2020 California campaign - "COVID"
- 2019 Texas Permian Basin campaign - "Permian"
- 2018 California campaign - "CalCH4"

These campaigns were selected for their diversity of surface conditions, spatial resolutions, and source types. Domain experts reviewed all methane matched filter product flightlines to filter out severe systematic artifacts, orthorectification artifacts, and erroneously identified methane source candidates.

Based on the resulting methane source coordinates, 256x256 pixel tiles were sampled with a per-flightline plume-to-background tile sampling ratio of 1:20 to create a classification dataset.

A15L-1395

Methodology

For **per-campaign models**, we train the GoogLeNet CNN classification model (Szegedy et al., 2014) on each campaign's sampled plume and background tiles. A test set is held out from the training set, on which all performance metrics are calculated. We also leverage Sharpness-Aware Minimization (Foret et al., 2021) and simple data augmentation for improved generalization and robustness.

For the **multi-campaign model**, we train the same model on tiles from all campaigns. To ensure that each campaign contributes to the model equally, some tiles from underrepresented campaigns are repeated each training epoch to match tile quantities in larger campaigns.

Saliency map generation typically requires expensive segmentation labels. Instead, by replacing the pooling and fully connected layers at the end of the model architecture with a 1x1 convolution layer and implementing shift-and-stitch interlacing (Long et al., 2015), we can **convert the classification CNN into a saliency map FCN**.

Results

The results table presents detection performance of the four models on all test datasets. The prototype model was trained using a set of ~200 plume tiles sampled from flightlines observed in early AVIRIS-NG methane survey campaigns in 2015-2016. Models constructed using the curated plumes and diverse background samples in the datasets described in this work consistently outperform the prototype model. Per-campaign models perform well within their original campaigns, but give mixed results when applied to other campaigns.
The multi-campaign model performs best overall, even outperforming some models on their original campaigns.

Figure 1 shows an example saliency map correctly highlighting methane plumes in a COVID campaign flightline. **Figure 2** breaks down true positive and false negative detections of plumes with IPCC source attributions. We do not identify any underperformance on specific IPCC sectors.

Acknowledgements
This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Copyright 2022 California Institute of Technology. U.S. Government sponsorship acknowledged.

Plume Classification on Multicampaign AVIRIS-NG Validation Data

Training Data + Model	Test Data: F1=2(pre+rec)/(pre*rec)				Precision = tp/(tp+fp)				Recall=tp/(tp+fn)			
	$CalCH_4$	Permian	COVID	Multi	$CalCH_4$	Permian	COVID	Multi	$CalCH_4$	Permian	COVID	Multi
Prototype	*17*	*23*	*15*	*19*	*10*	*15*	*9*	*11*	*54*	*50*	86	*57*
$CalCH_4$	73	52	77	61	79	43	65	52	69	68	95	73
Permian	62	**75**	50	67	57	63	39	56	67	**94**	*73*	84
COVID	79	62	**82**	70	87	59	**73**	67	72	66	95	72
Multi	**85**	73	78	**77**	**88**	**64**	65	**68**	**81**	86	**97**	**87**

Train =Test Campaign **Best score** *Worst score*

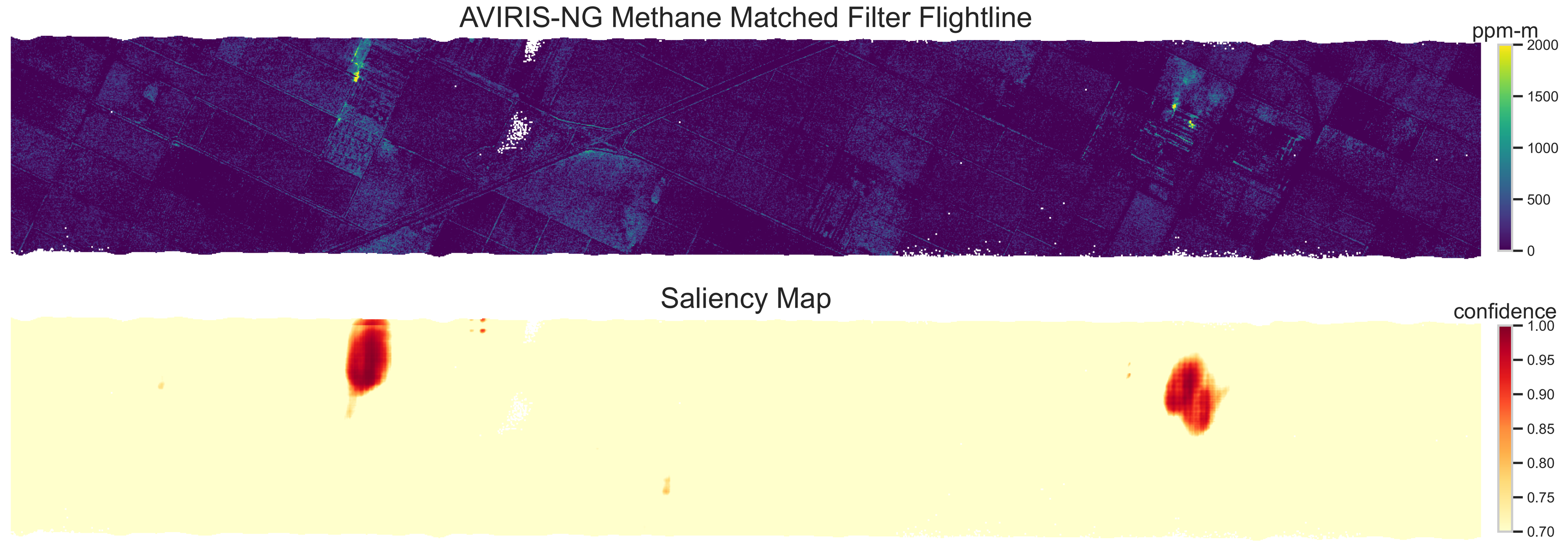

Fig 1. Example saliency map output by the multi-campaign model.

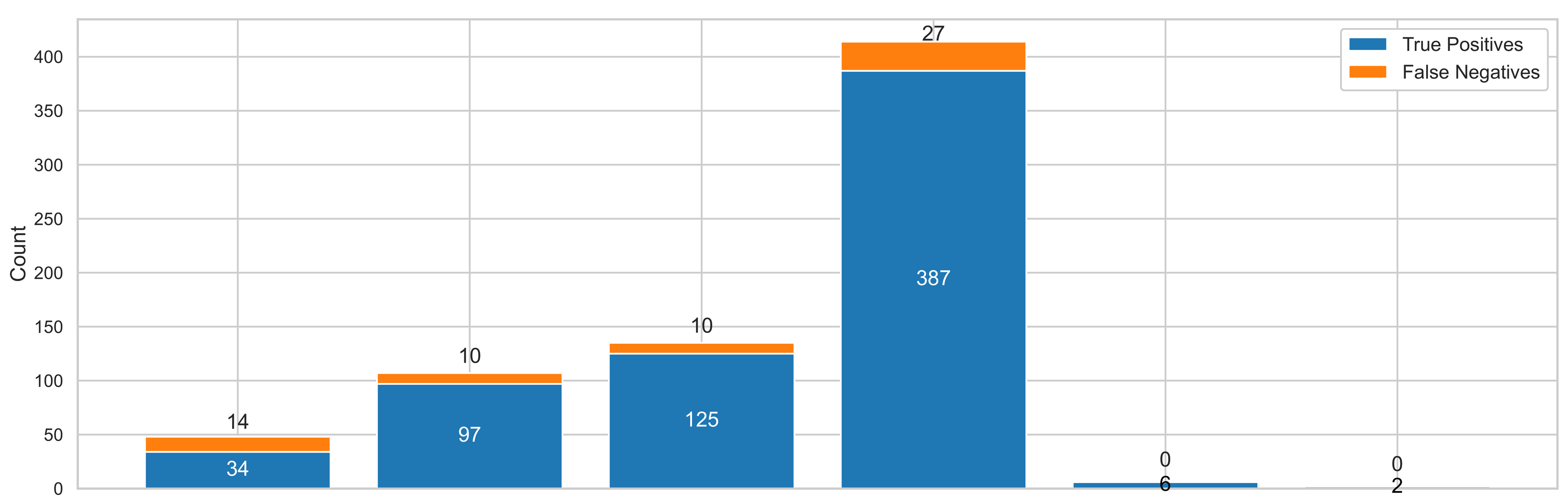

Fig 2. True positive and false negative plume detections by IPCC sector.