Deep Learning Saliency and Segmentation Methods for Robust Methane Plume Detection

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

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

Introduction

Identification of global methane sources is critical for the quantification and mitigation of this greenhouse gas. Future missions such as Carbon Mapper will provide global imaging spectroscopy observations, and an automated plume detection pipeline is needed for the timely mapping of methane sources.

We improve upon our prior work and contribute the following:

- **We curate a multicampaign methane matched filter tiled dataset.**
- **We develop multiple deep learning saliency and segmentation models.**
- **We evaluate and report tilewise classification and pixelwise segmentation performances of these models.**
- **We visualize detections on entire flightlines.**

Dataset

We combine flightlines from four campaigns. Train and test sets are spatially stratified by flightline to avoid leakage, as shown in Figure 1.
Tiles of size (256, 256) are then sampled at a ratio that simulates a realistic flight campaign, where plumes are much rarer than the background. Table 1 shows the number of tiles for each dataset from each campaign.

Labeled plume sources for classification were reviewed by Subject Matter Experts to minimize artifacts, systematics, and other erroneous labels.
Labeled plume masks for segmentation were generated algorithmically. While not perfectly accurate, it is sufficient for our detection task. Some error in the predicted mask is acceptable as long as the plume is detected in the first place.

Instrument	Campaign	Year	Train Plumes	Train BGs	Test Plumes	Test BGs
AVIRIS-NG	"CACH4"	2018	183	3,149	102	1,278
	"Permian"	2019	1,074	10,010	125	2,421
	"COVID"	2020	146	3,283	70	1,902
GAO	"CARB"	2020	87	2,308	27	542
Combined	**"Multicampaign"**		**1,490**	**18,750**	**324**	**6,143**

Table 1. *Number of plume and background tiles from each campaign.*

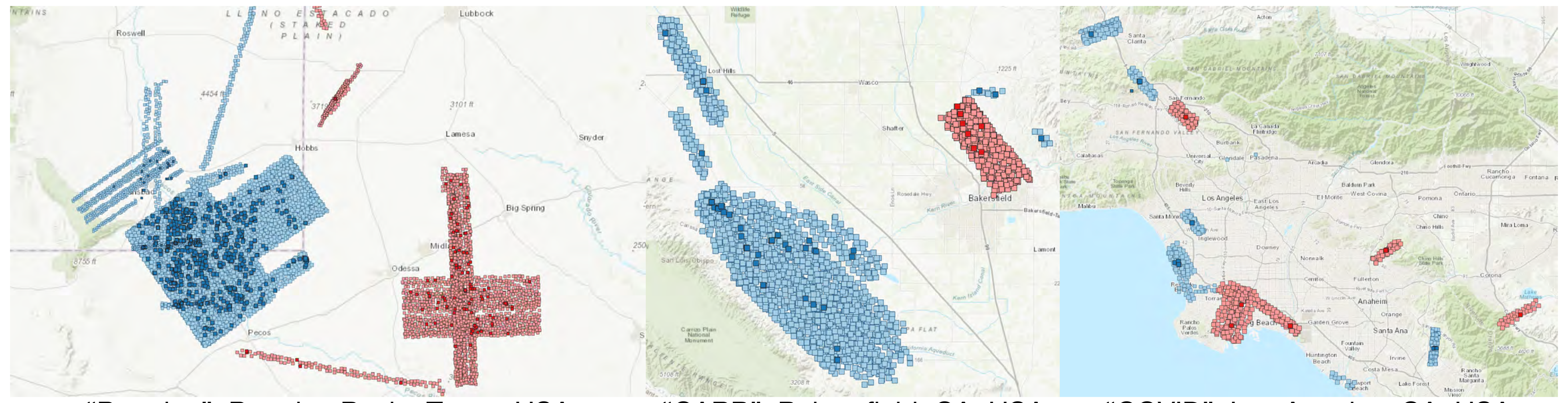


Figure 1. *Maps of some of the tiles sampled for model training. Blue tiles are in the training set, and red tiles are in the test set. Darker tiles contain labeled plumes.*

Methods

We demonstrate four models on the multicampaign dataset:

- **GoogLeNet CNN** classification model (Szegedy et al., CVPR 2014)
- **GoogLeNet FCN** segmentation model converted from the CNN with shift-and-stitch (Long et al., CVPR 2015)
- **U-Net** segmentation model trained with both classification and segmentation losses (BCE+FocalTversky) (Ronneberger et al., 2015)
- **UPerNet** object branch segmentation model using the pretrained GoogLeNet CNN above as a backbone, shown in Figure 2 (Xiao et al., ECCV 2018)

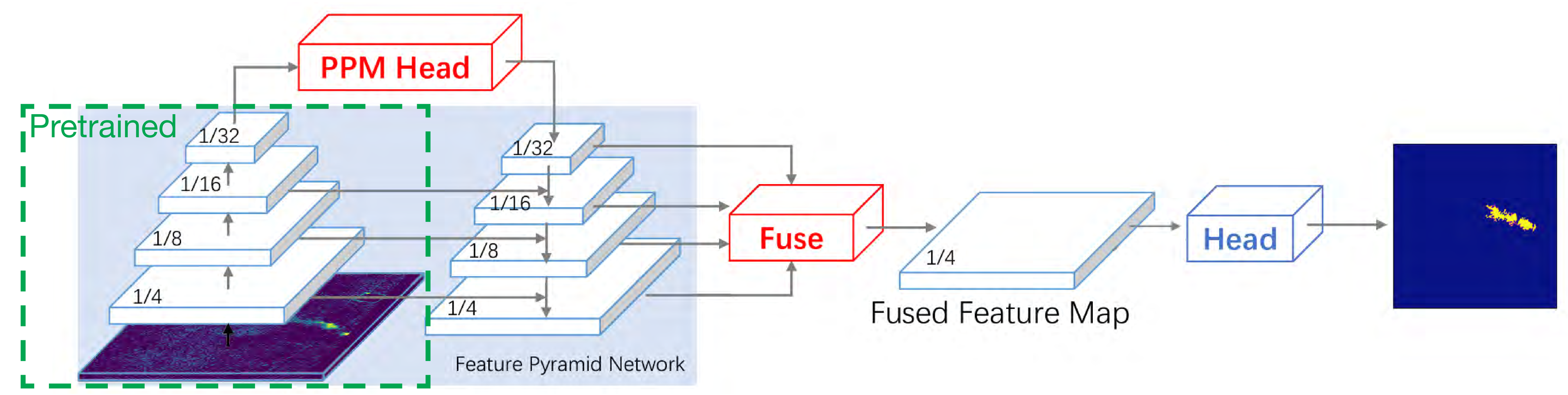


Figure 2. *A diagram of the UPerNet model, modified from Xiao et al. The model in the green box is GoogLeNet, pretrained on tile classification. It is frozen while the rest of the model is trained for plume segmentation.*

Results

We care about both detecting plume presence and generating decent masks. Therefore, we consider both of the following metrics:

- **Tilewise Classification F1 Score:** How well can the model classify whether a tile contains a plume? The maximum value of each predicted mask is considered the tilewise predicted value. The metric is calculated across all tiles in the test set.
- **Pixelwise Segmentation F1 Score:** How well can the model segment whether a pixel is part of a plume? The metric is calculated across all pixels in the test set.

Model	Classification			Segmentation		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
GoogLeNet	0.769	0.861	0.812	0.430*	0.505*	0.465*
U-Net	0.636	0.611	0.622	0.510	0.567	0.537
UPerNet	0.823	0.830	**0.826**	0.650	0.589	**0.618**

*Via Shift and Stitch FCN

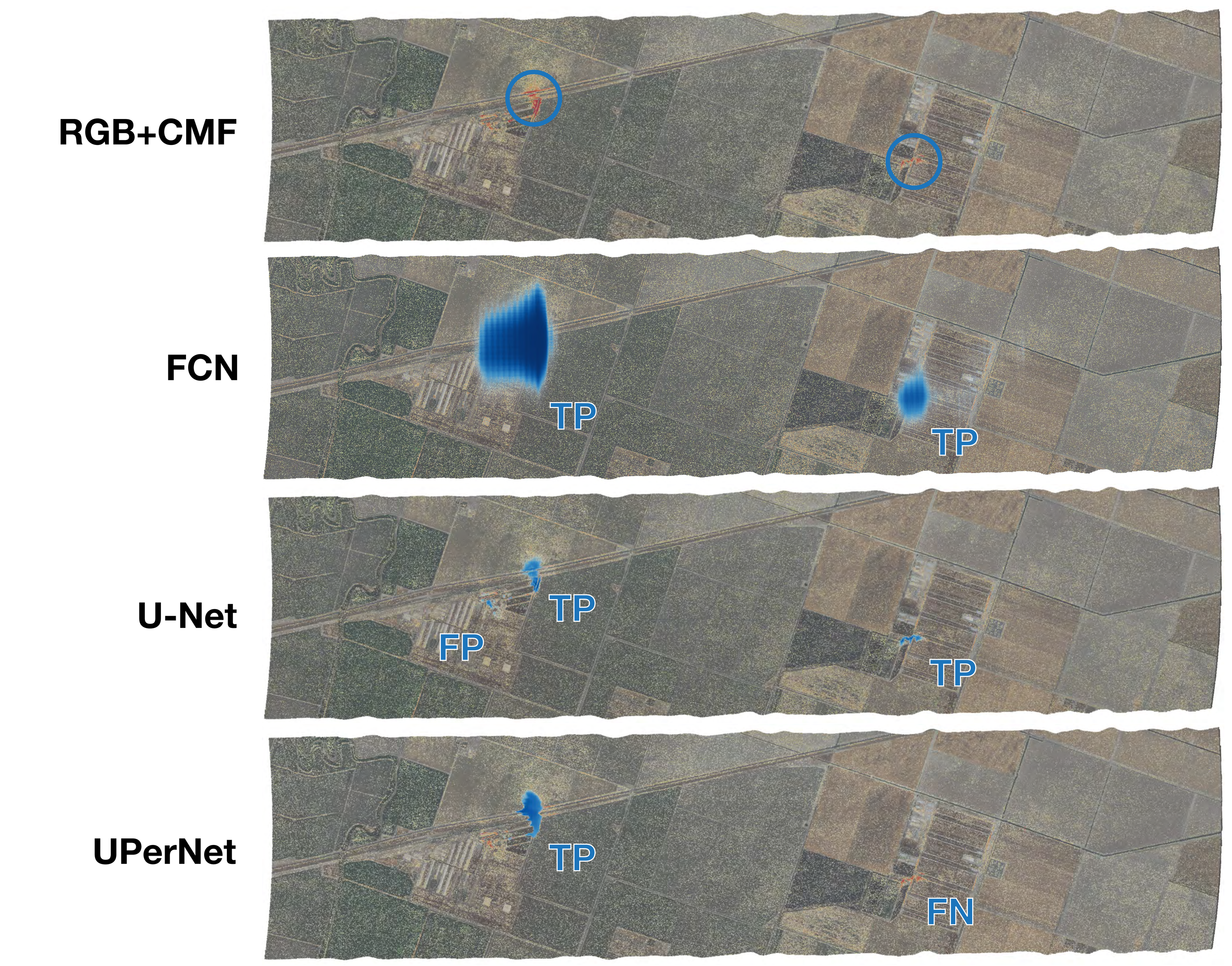
Table 2. *Number of plume and background tiles from each campaign.*

The UPerNet outperforms all other models on both metrics. Starting with a model backbone that can already distinguish between plumes and artifacts/noise provides a big benefit to classification, especially false positive reduction, while the rest of the architecture (which shares the same design principles as the U-Net) is able to outperform the U-Net in segmentation. In implementation, it is also faster to train the CNN then the UPerNet than it is to train the U-Net from scratch.

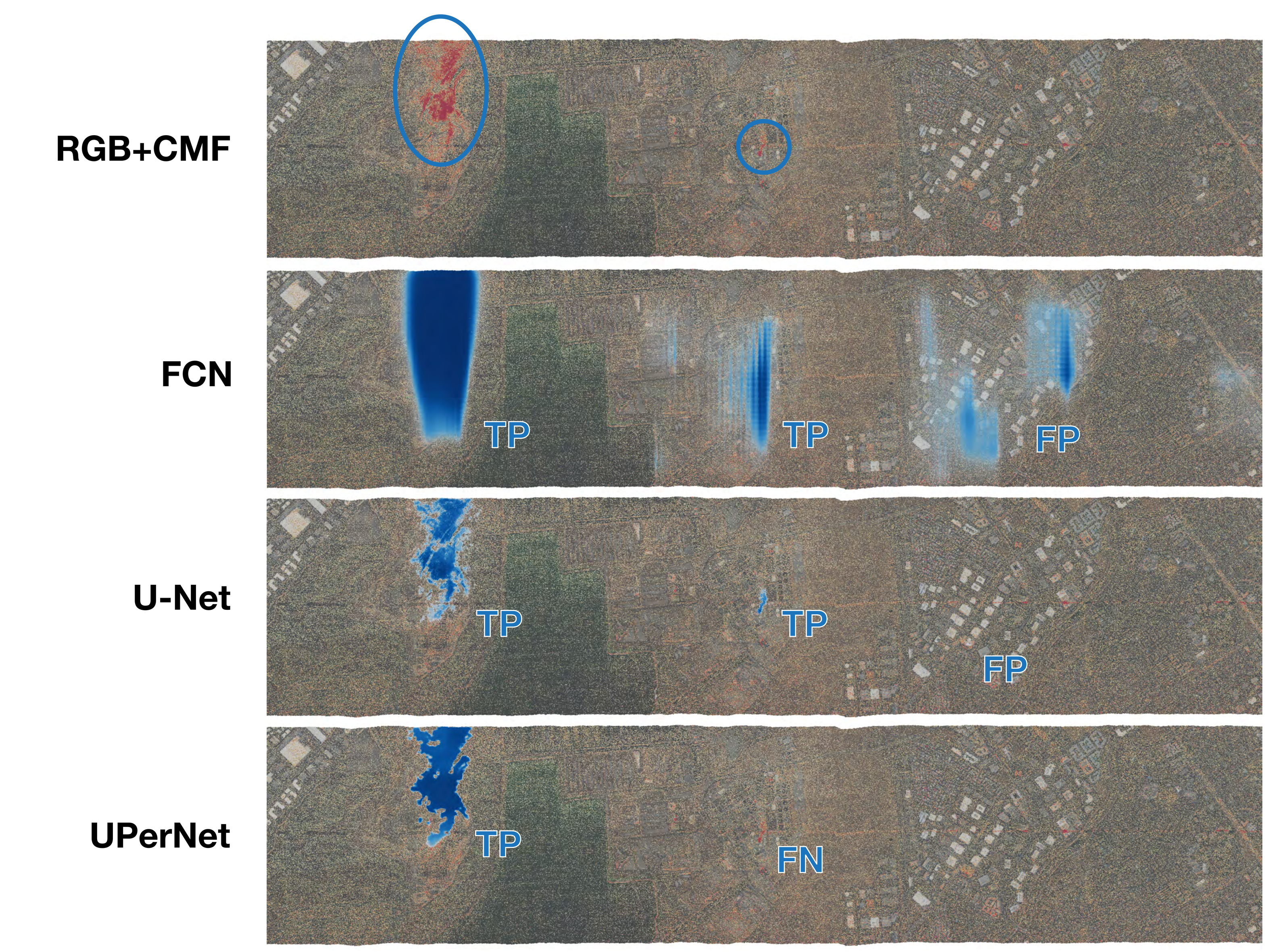
Acknowledgements

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Visualizations



ang20200724t204341_ch4mf_v2y1_img



ang20180927t200251_cmf_v2t1_img

A43R-3011