MethaneMapper: Spectral Absorption aware Hyperspectral Transformer for Methane Detection

Satish Kumar satishkumar@ucsb.edu
Ivan Arevalo ifa@ucsb.edu
ASM Iftekhar iftekhar@ucsb.edu
B S Manjunath manj@ucsb.edu

Department of Electrical and Computer Engineering
University of California Santa Barbara

Abstract

Methane (CH_4) is the chief contributor to global climate change. Recent Airborne Visible-Infrared Imaging Spectrometer-Next Generation (AVIRIS-NG) has been very useful in quantitative mapping of methane emissions. Existing methods for analyzing this data are sensitive to local terrain conditions, often require manual inspection from domain experts, prone to significant error and hence are not scalable. To address these challenges, we propose a novel end-to-end spectral absorption wavelength aware transformer network, MethaneMapper, to detect and quantify the emissions. MethaneMapper introduces two novel modules that help to locate the most relevant methane plume regions in the spectral domain and uses them to localize these accurately. Thorough evaluation shows that MethaneMapper achieves 0.63 mAP in detection and reduces the model size (by 5×) compared to the current state of the art. In addition, we also introduce a large-scale dataset of methane plume segmentation mask for over 1200 AVIRIS-NG flight lines from 2015-2022. It contains over 4000 methane plume sites. Our dataset will provide researchers the opportunity to develop and advance new methods for tackling this challenging green-house gas detection problem with significant broader social impact. Dataset and source code link[1].

1. Introduction

We consider the problem of detecting and localizing methane (CH_4) plumes from hyperspectral imaging data. Detecting and localizing potential CH_4 hot spots is a necessary first step in combating global warming due to greenhouse gas emissions. Methane gas is estimated to contribute 20% of global warming induced by greenhouse gasses [24] with a Global Warming Potential (GWP) 86 times higher than carbon dioxide (CO_2) in a 20 year period [33]. To put into perspective, the amount of environmental damage that CO_2 can do in 100 years, CH_4 can do in 1.2 years. Hence it is critical to monitor and curb the CH_4 emissions. While CH_4 emission has many sources, of particular interest are those from oil and natural gas industries. According to the United States Environmental Protection Agency report, CH_4 emissions from these industries accounts to 84 million tons per year [18]. These CH_4 emissions emanate from specific locations, mainly from pipeline leakages, storage tank leak or leakage from oil extraction point.

Current efforts to detect these sources mostly depend on aerial imagery. The Jet Propulsion Laboratory (JPL) has conducted thousands of aerial surveys in the last decade to collect data using an airborne sensor AVIRIS-NG [21]. Several methods have been proposed to detect potential emission sites from such imagery, for example, see [8, 9, 35, 38, 40, 41]. However, these methods are in general very sensitive to background context and land-cover types, resulting in a large number of false positives that often require significant domain expert time to correct the detections. The primary reason is that these pixel-based methods are solely dependent on spectral correlations for detection. Spatial information can be very effective in reducing these false positives as CH_4 plumes exhibit a plume-like structure morphology. There has been recent efforts in utilizing spatial correlation using deep learning methods [22, 29], however, these works don't leverage spectral properties to filter out confusers. For example, methane has similar spectral properties as white-painted commercial roofs or paved surfaces such as airport asphalts [1]. This paper presents a novel deep-network based solution to minimize the effects of such confusers in accurately localizing methane plumes.

Our proposed approach, referred to as the MethaneMapper (MM), adapts the DETR [4], a transformer model that combines the spectral and spatial correlations in the imaging data to generate a map of potential methane (CH_4) plume candidates. These candidates reduce the search space for a hyperspectral decoder to detect CH_4 plumes and re-

[1] *https://github.com/UCSB-VRL/MethaneMapper-Spectral-Absorption-aware-Hyperspectral-Transformer-for-Methane-Detection*

move potential confusers. MM is a light-weight end-to-end single-stage CH_4 detector and introduces two novel modules: a *Spectral Feature Generator* and a *Query Refiner*. The former generates spectral features from a linear filter that maximizes the CH_4-to-noise ratio in the presence of additive background noise, while the latter integrates these features for decoding.

A major bottle neck for development of CH_4 detection methods is the limited availability of public training data. To address this, another significant contribution of this research is the introduction of a new Methane Hot Spots (MHS) dataset, largest of its kind available for computer vision researchers. MHS is curated by systematically collecting information from different publicly available datasets (airborne sensor [6], Non-profits [3, 31] and satellites [34]) and generating the annotations as described in Section 4.1.1. This curated dataset contains methane segmentation masks for over 1200 AVIRIS-NG flight lines from years 2015 to 2022. Each flight line contains anywhere from 3-4 CH_4 plume sites for a total of 4000 in the MHS dataset.

Our contributions can be summarized as follows:

1. We introduce a novel single-stage end-to-end approach for methane plume detection using a hyperspectral transformer. The two modules, *Spectral Feature Generator* and *Query Refiner*, work together to improve upon the traditional transformer design and enable localization of potential methane hot spots in the hyperspectral images using a Spectral-Aware Linear Filter and refine the query representation for better decoding.

2. A new *Spectral Linear Filter (SLF)* improves upon traditional linear filters by strategically picking correlated pixels in spectal domain to better whiten background distribution and amplify methane signal.

3. A new benchmark dataset, MHS, provides the largest ($\sim$ 35$\times$) publicly available dataset of annotated AVIRIS-NG flight lines from years 2015-2022.

2. Related Works

Our work is at the intersection of hyperspectral data for CH_4 detection, deterministic linear filtering methods for spectral features and encoder-decoder based transformer. A review of the pertinent related works is given below.

There are several recent papers on detecting methane plumes from the airborne imaging spectrometer AVIRIS-NG [21]. This includes the Iterative Maximum a Posterior Differential Optical Absorption Spectroscopy algorithm (IMAP-DOAS) [7, 8] and matched filters [9, 35, 38, 40, 41]. IMAP-DOAS requires data from 2 hyperspectral sensors, one airborne and another on ground, hence not very practical for most application scenarios. Matched-filter based methods use background statistics to normalize

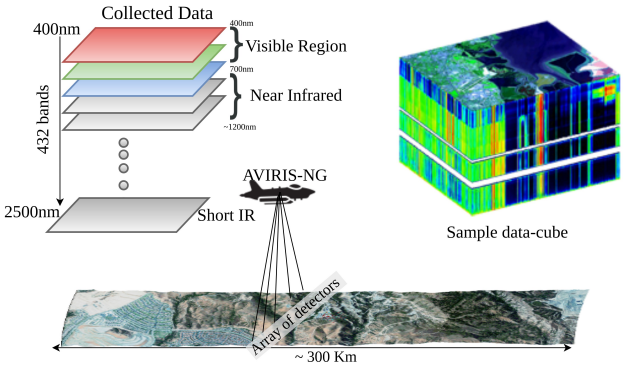

Figure 1. *Depiction of data collection process. Each flightline is* $\sim$ 300 *km long. An array of 598 sensors records data at 1.5m/pixel spatial resolution. All flightlines are ortho-corrected. Each datacube is of dimension* $\sim 25000\times \sim 1500 \times 432$.

the spectral signals and match with the CH_4 spectral signature at every spatial location (pixel-wise). This process, however, is sensitive to surface albedo and land cover with spectral absorption similar to CH_4, leading to spurious detections. Domain experts must then manually inspect each flight line to identify and delineate real CH_4 plumes [39]. To suppress the effect of false positives due to variability of elements on ground, *Christopher et.al.* [10] and *Thorbe et. al.* [41] introduced cluster-tuned matched filter. It involves clustering the pixels with similar spectral properties using k-means clustering. Both IMAP-DOAS and all versions of matched filters are heavily prone to false positives as the information is processed pixel-wise.

Machine learning approaches have been used for target detection, including CH_4 identification, in hyperspectral imagery [12, 17, 28, 36, 37]. Similar to matched-filtering, these methods do not take into account the influence of confusers on the CH_4 spectral signature and have similar issues concerning false positives. Recently introduced deep learning based H-mrcnn model [29] focus on capturing spatial correlation. H-mrcnn is an ensemble of mask-rcnncite networks processing blocks of hyperspectral data. This block processing in the spectral domain is inefficient and often results in overall poor performance. Methanet [22] is a more recent work focusing on estimating methane concentration from matched-filter data. In this regard, our proposed MethaneMapper uses both spectral and spatial correlation to accurately delineates CH_4 plumes.

Datasets: The only dataset publicly available with annotation for CH_4 plume detection is JPL-CH_4-detection2017-V1.0 dataset [39]. It contains only 46 AVIRIS-NG [21] flight lines in the US Four-Corners region. Deep learning architectures require a large number of annotated samples, and for this reason we introduce the new MHS dataset with over 1200 annotated flightlines and $\sim$ 4000 plume sites.

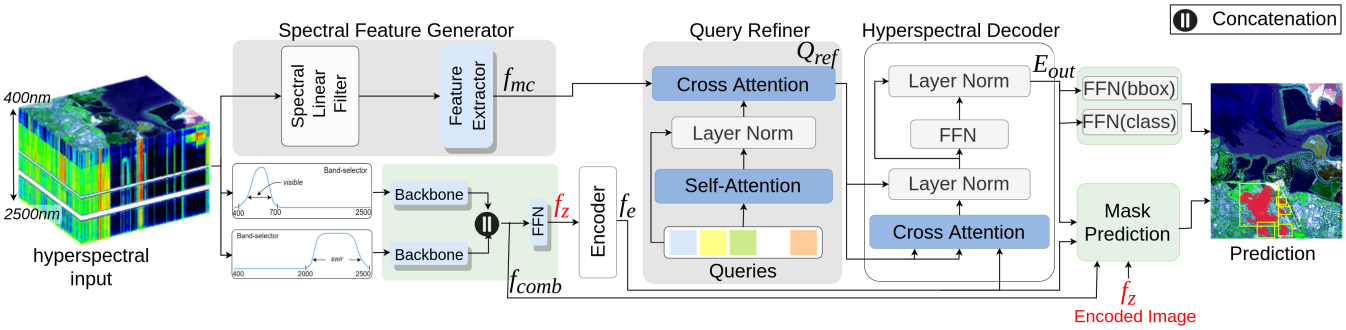

Figure 2. *Overview of MethaneMapper (MM) architecture. Given a hyperspectral image, our RGB ($400nm - 700nm$) and SWIR ($2000nm - 2500nm$) band-pass filters passes a subset of channels in desired wavelength range and feed them to CNN backbones (ResNet) to extract features. These features are concatenated and fed to Transformer Encoder. Parallelly, our Spectral Feature Generator (SFG) modules takes in all channels of input image and generate methane candidates features. Next these candidates are sent to Query Refiner (QR) to refine queries. Then these queries decoded using encoded feature from Transformer Encoder. Finally each decoded query is used to predict a plume mask via Mask Prediction and, bounding box and class via FFNs (Feed Forward Network).*

3. MethaneMapper (MM) Architecture

3.1. Data Overview

AVIRIS-NG hyperspectral imaging sensors capture spectral radiance values from N_0 ($N_0 = 432$) channels corresponding to wavelengths ranging from $400nm - 2500nm$ as shown in Fig. 1. The complete hyperspectral image is represented as $\mathbf{x} \in \mathbb{R}^{H_0 \times W_0 \times N_0}$ where H_0, W_0 are the height & width, respectively, and $N_0 = 432$ is number of channels. This hyperspectral data includes a very weak signature of CH_4 around 2100-2400nm, conflated with radiations from the surrounding land cover and background clutter. A single flight-line could be over a couple miles long (about 25K pixels in one of the dimensions), with an array of sensors recording the data at 1.5m/pixel resolution. The images are orthorectified before processing.

3.2. Technical Overview

Referring to Fig.2, MM contains the following main components: (i) 2 CNN backbones to extract a compact feature representation of the spectral regions of interest from the hyperspectral image, (ii) a *Spectral Feature Generator* (**SFG**), and (iii) a *Query Refiner* (**QR**) in between an encoder-decoder pair (inspired by GTNet [20], SSRT [19]). The hyperspectral image is first processed through two separate band-pass filters to select the channels in visible ($400 - 700nm$) and short-wave infrared (SWIR)($2000 - 2500nm$) wavelength regions, and are then passed through CNN backbones. Output of these backbones are concatenated together and then encoded using a transformer encoder.

The **SFG** (Sec. 3.4) takes in all channels of the hyperspectral image and process them through a spectral linear filter. The **SFG** exploits the spectral correlation to generate methane candidates feature maps and passes them to **QR**. The **QR** (Sec. 3.5) uses these methane candidates to refine the learnable queries. Our hyperspectral decoder takes the encoded features from the encoder and refined queries from **QR** to generate the embeddings. The mask-prediction layer processes these embeddings along with the feature pyramid from the backbone layers to generate the final methane-plume segmentation prediction.

These individual blocks are discussed in more detail below.

3.3. Bandpass filtering for the Encoder

The HSI is processed by two parallel band-pass filters; a visible wavelength ($400 - 700nm$) (RGB) and a short-wave infrared wavelength ($2000 - 2500nm$) (SWIR) band-pass filter. The RGB filter results in a 3 channel output corresponding to the normal red, green, and blue wavelengths. The SWIR generates channels, approximately 5nm apart. The filtered outputs are $\mathbf{x}_{rgb} \in \mathbb{R}^{H_0 \times W_0 \times 3}$ and $\mathbf{x}_{swir} \in \mathbb{R}^{H_0 \times W_0 \times 100}$. Using $\mathbf{x}_{rgb}$ and $\mathbf{x}_{swir}$, two conventional CNN backbones (e.g. ResNet-50 [16, 25]) generate two feature maps respectively of size $\in \mathbb{R}^{H \times W \times N}$. Here $H = \frac{H_0}{32}$, $W = \frac{W_0}{32}$ and $N = 2048$ typically. We concatenate these feature maps along channel dimension and project through a 1×1 convolution layer to retain channel dimension of N. The resulting output is $f_{comb} \in \mathbb{R}^{H \times W \times N}$.

Following the standard architecture of transformer encoder from previous works [4, 19, 20, 26, 43], we reduce the channel dimension of f_{comb} using 1×1 convolution to $f_z \in \mathbb{R}^{H \times W \times d}$ and supplement position information by adding a fixed positional embedding $p \in \mathbb{R}^{H \times W \times d}$. The encoder consists of a stack of multi-head self-attention modules and feed-forward networks (FFN). The encoded feature map is $f_e \in \mathbb{R}^{H \times W \times d}$:

$$f_e = \text{Encoder}(f_z, p) \tag{1}$$

3.4. Spectral Feature Generator (SFG)

In parallel, the input hyperspectral image is processed by the **SFG** module to generate methane candidates feature map f_{mc}, providing the **QR** module with spatial information to help the network delineate the methane plumes.

The **SFG** consist of a spectral linear filter (SLF) and a Feature Extractor (e.g.ResNet-50 [16]). The most common linear filtering approach for detecting CH_4 is to take each pixel from the input hyperspectral image $\{\mathbf{x}_{ij} \mid \mathbf{x}_{ij} \in \mathbb{R}^{1\times1\times N_0}\}_{i,j=1}^{H_0,W_0}$ and project it onto a CH_4 spectral absorption signature vector of same size [13]. This is to reduce the interference from ground terrain and amplify the CH_4 visibility in that pixel. Accurately modeling SLF is critical given that it is designed to reduce ground terrain interference.To model **SLF** we use the most common approach to matched filtering from information theory [42].

Spectral Linear Filter (SLF): The design of SLF is dependent on the spectral absorption pattern of CH_4 gas [13] and distribution of ground terrain. Since our signal of interest, CH_4, is very weak, traditional methods of linear filtering [40, 41] are not effective. The conventional methods to whiten the ground terrain noise includes calculating the covariance ($\mathbf{Cov} \in \mathbb{R}^{N_0\times N_0}$) of background by selecting a set of 10-15 adjacent columns $\{\mathbf{x}_i \mid \mathbf{x}_i \in \mathbb{R}^{1\times H_0\times N_0}\}_{i=1}^{W_0}$). However, in a given flight-line, the terrain changes frequently, from water bodies to bare soil, vegetation, buildings and other urban structures. Therefore single approximation of the covariance can not provide correct estimate of CH_4 and a localized context-based whitening will be more effective. To address this problem, we took a very simple and effective approach of doing land cover classification and segmentation [11, 32, 44], and then compute covariance per class from the land cover. More details in supplementary materials. This improves the quality of methane candidates in presence of confusers (materials with similar spectral absorption patterns as CH_4) and also in cases where CH_4 concentration is low. The final **SLF** design with per class covariance is:

$$\mathbf{SLF}(\mathbf{x}_{ij}) = \frac{(\mathbf{x}_{ij}-\mu_k)^T\mathbf{Cov}_k^{-1}t}{\sqrt{t^T\mathbf{Cov}_k^{-1}t}} \;\forall\; (i,j) \in \text{class } k \quad (2)$$

where t represents the spectral absorption pattern [13] of CH_4 gas, and $\mathbf{Cov}_k$, μ_k are the covariance and mean of k^{th} class respectively. $\mathbf{x}_{ij}$ represents the pixel in input hyperspectral image at (i,j) index in k^{th} class. This operation generates a 2-D spatial CH_4 candidates map of size $\mathbb{R}^{H_0\times W_0}$. Next this CH_4 candidates map is fed to a Feature Extractor to generate CH_4 candidates feature map f_{mc}. Details of the land cover segmentation/classification and complete SLF derivation are in the Supplementary materials.

$$f_{mc} = \text{FeatureExtractor}(\ \mathbf{SLF}(\mathbf{x}_{ij})\ \forall\ i,j) \quad (3)$$

3.5. Query Refiner (QR)

Next the methane candidate feature map $f_{mc} \in \mathbb{R}^{H\times W\times d}$ is fed to the **QR** module along with a set of 100 learnable queries $Q \in \mathbb{R}^{100\times d}$. The f_{mc} refines the learnable queries via cross-attention mechanism. This operation provides a narrow search space for the queries. The **QR** module follows a transformer decoder-like architecture inspired from [19,20]. The randomly initialized queries $Q \in \mathbb{R}^{100\times d}$ are first passed through a self-attention layer to attend to themselves. Next, these queries attend to our methane candidates feature map f_{mc} from **SFG** module through a cross-attention layer. The methane candidates feature map serves as key-values pairs in our attention architecture. The output of **QR** is Q_{ref}.

$$Q_{ref} = \mathbf{QR}(f_{mc}, Q) \quad (4)$$

3.6. Hyperspectal Decoder

The Q_{ref} is fed to the decoder module along with encoder output f_e to generate output embeddings. Our hyperspectral decoder follows the standard architecture with a minor difference. There are no self-attention layers, just stack of multi-headed cross attention layers. The refined queries are transformed into output embeddings $E_{out} \in \mathbb{R}^{100\times d}$.

$$E_{out} = \text{Decoder}(f_e, p, Q_{ref}) \quad (5)$$

3.7. Box and Mask Prediction

The decoder output embeddings (E_{out}) are fed to two Feed Forward Network (FFNs) and a Mask prediction layer. The outputs of the FFNs are the bounding boxes covering each CH_4 plume and a confidence score corresponding to each box. The mask-prediction module follows the standard segmentation head of DETR [4]. It computes multi-head attention scores of each embedding over the f_e (Eq. 1), generating a low-resolution heatmap for each embedding. To make the final prediction a Feature Pyramid Network [14] like structure is used. Each heatmap is designed to capture one methane plume. A simple thresholding is used to merge the heatmaps as final segmentation mask.

$$mask = \text{Mask\_pred}(E_{out}, f_e, f_{comb}) \quad (6)$$

3.8. Training and Inference

We train MethaneMapper in two stages; first we train bounding box detection corresponding to each CH_4 plume, and second by freezing the box detection network and training only the mask prediction module. We also trained both box and mask prediction modules end-to-end and achieved similar performance. We use a similar two-stage loss strategy for training MethaneMapper as that used in DETR [4]:

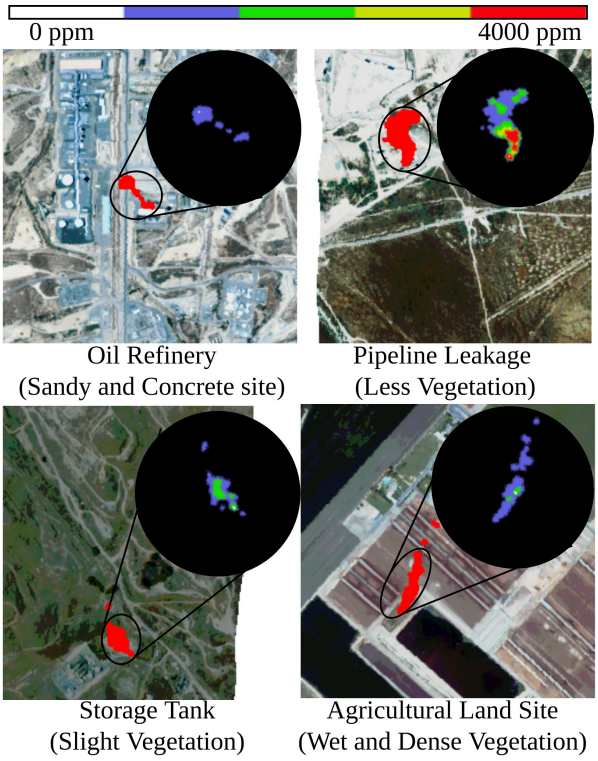

Figure 3. *Sample images from MHS dataset. The colormap in black circle shows concentration maps corresponding to the plume mask shown in red. We are showing different types of leakage sources and land cover types. For better visualization, we plotted the binary mask on color image created using visible bands of hyperspectral image.*

first stage is the bipartite matching between the predictions and the ground truths both in bounding box and mask prediction, and then second stage is loss calculation for the matched pairs. The bipartite matching employs the Hungarian algorithm [4] to find the optimal matching between the predictions and the ground truths. After this matching, every prediction is associated with a ground truth. Next, we calculate the l_1 and $GIoU$ loss on both box and mask predictions and cross entropy loss for class prediction [4].

Inference: The inference pipeline is similar to training pipeline and can be implemented using approximately 50 lines of code. During inference, we first filter the detections with confidences below 50% and a per-pixel max to determine which pixels are predicted to belong to a CH4 plume.

4. Methane Hot Spots (MHS) dataset

Another significant contribution of this work is a large scale curated MHS dataset. It contains the AVIRIS-NG spectral data with wavelength ranging from $380nm$ to $2510nm$, a $5nm$ sampling [21], and capturing 432 channels per pixel. The images from the flight-line are orthorectified and of size $\sim 23K\times \sim 1.5K \times 432$. The only currently publicly-available dataset with methane plume segmentation masks is the JPL-CH4-detection-V1.0 [39] dataset released by JPL-NASA in 2017. The MHS dataset has approximately 4000 plume sites corresponding to approximately 1200 AVIRIS-NG flightlines as shown in Table 1. MHS also has higher diversity data with flight lines spanning from 2015-2022 and covering terrain from 6 states– California, Nevada, New Mexico, Colorado, Midland Texas, and Virginia.

Dataset	MHS (Ours) Dataset	JPL-CH4 detection-V1.0 [39]
# plume sites	**3961**	161
# flightlines	**1185**	46
# point source	**3675**	114
# diffused source	**286**	57
Time period	**2015 - 2022 (8 years)**	2015 (1 year)
Segmentation Mask	**Yes**	**Yes**
Bonding box	**Yes**	No
Concentration map	**Yes**	No
Number of Regions	**6**	1

Table 1. *Statistics shows MHS dataset comparison with JPL-CH4-detection-V1.0 [39] dataset. Each flightline have multiple large and small plume sites. Each flightline have atleast 4 plume sites. The Point Source represents high concentration ($300kg/hr$) to leakage from sources like pipeline leak, storage tanks, oil and gas refineries. Diffused Source represent low concentration leakages from sources like biomass degradation in landfills. Our dataset is covers more diverse type of terrain over 6 states.*

Data Pruning: We selected AVIRIS-NG flight lines over varying regions as it covers a wide variety of CH_4 plume sources, such as leaks in oil and gas refineries, oil and gas extraction points, natural seeps, leaking underground storage tank, coal mines, dairy farms, landfill sites, and pipeline leaks. Along with varying emission sources, we selected regions with different types of ground terrains like, bare soil, rocks, mountains, light vegetation, water bodies and dense vegetation as shown with few samples in Fig. 3. Different types of ground terrain exhibit widely varying albedo and thus have a major impact on the quality of CH_4 detections as shown in Fig. 4. Given this, training models with diverse ground terrain data leads to a more robust model.

4.1. Concentration map and Segmentation mask

Concentration map is provided in the form of a matrix of spatial dimensions same as the flightline ($\sim 23k\times \sim 1.5k \times 1$). There is one concentration map per flight-line (orthorectified). It shows methane concentration in parts-per-million (ppm) per-pixel on the ground. Pixel-regions with no methane presence are set to zero.

Segmentation mask provided in the format of a *"png"* image file with three channels and of the same spatial dimension as the corresponding flight line ($\sim 23k \times \sim 1.5k \times 3$). The segmentation mask is obtained from the concentration mask file by setting all pixel values above zero to represent methane plumes. We manually annotated *Point Source* and *Diffused Source* based on the type of ground terrain and concentration of methane gas. Following the benchmark dataset [39], three channels are used to color code *Point Source* (Red) and *Diffused Source* (Green). The distinction of *Point Source* and *Diffused Source* is derived from the JPL-CH4-detection-V1.0 benchmark dataset [39]. Our annotation style is also consistent with the JPL-CH4-detection-V1.0 benchmark dataset [39], so that both datasets can be merged seamlessly.

4.1.1 Constructing Concentration map

Concentration maps are generated by mapping expert-annotated methane-plume concentration maps to the ortho-corrected AVIRIS-NG flightlines. These methane plume annotations are systematically collected from a non-profit [3] entity. They provide concentration masks of methane emissions in 150×150 size patches along with location information from different sources (airborne sensors [6], satellites [34]). In order to map these patches from different sources to the AVIRIS-NG flight-lines, we use the pixel coordinate locations provided for both the annotations and flight-lines. We use this information to create a homography transformation to map each pixel to its corresponding location in the flight-line. Fig. 3 shows a sample of varying types of terrains with CH_4 segmentation mask in red and concentration mask in black circle. Details about matching the resolution, ortho-correction, and transformation are discussed in supplementary materials. The patch annotations are verified by experts visiting the physical location of emission the same day [2]. Most of the regions in California are verified by physical visits by California Air Resource Board [2, 3].

4.2. MHS Statistics

MHS statistics and properties are summarized in Table 1.

Annotations: MHS provides both segmentation masks and concentration maps which enable development of deep learning algorithms than can produce both CH_4 plume location and concentration predictions.

Diversity: MHS dataset includes AVARIS-NG flightlines spanning 8 years (2015 - 2022) from six states in the U.S.: California, Nevada, New Mexico, Colorado, Texas, and Virginia.

Data Split: We divide MHS dataset into train/test splits of 80-20% with overlapping time periods and locations. Our dataset covers 6 states. Each state has sub-regions/locations (e.g. Permian basin) that are covered by multiple non-overlapping flightlines ($25k \times 1.5k \times 432$ pixels). These flightlines are split into train and test sets. In each set, we create patches ($256 \times 256 \times 432$ pixels) from the corresponding flightlines. From the patches/tiles, we take all positives patches (methane (CH_4)) and randomly sample equal number of negative (no-CH_4) patches. This is done for both train and test sets separately to balance the data and we refer to Section 6.2 for detailed ablation studies.

5. Experimental settings

Evaluation Metrics: Following the evaluation protocol of H-mrcnn [29] we report our performance in mean intersection-over-union (mIOU). Here, mIOU indicates the overlap between the predicted and the ground truth CH_4 plume masks. ED represents the accuracy in plume core prediction. Additionally, as first stage of our two stage training procedure contains bounding box prediction, we also report our performance in predicting plume bounding boxes in terms of mean Average Precision (mAP) which tells us the effectiveness of MethaneMapper in eliminating the false positives in plume prediction.

Data Pre-Processing: Each input hyperspectral image is approximately of size $25000 \times 1500 \times 432$ taking up memory space of $55 - 60$ GB. We create tiles of each image in spatial domain, each tile is of size $256 \times 256 \times 432$ [29] with an overlap of 128. The CH_4 plume is available in very few pixels in the whole image, 90% of the tiles are negative samples (no methane, just ground terrain). We can not use the whole hyperspectral image because of GPU memory limitations

Implementation Details: The band-selectors module takes 432-channels hyperspectral image as input, the RGB band-selector picks 60 channel from $400nm - 700nm$ wavelength range and creates a 3-channel RGB image, the SWIR band-selector picks 100 channel from wavelength range $2000nm - 2500nm$. These input images are passed to two ResNet-50 [16] feature extractor backbones. The backbone networks are initialized with DETR [4] trained on COCO dataset [30] and input layer initialized randomly [15]. The transformer encoder-decoder and our query refiner have 6 layers and 8 heads. We initialized the transformer encoder-decoder with weights extracted and stripped from DETR [4] model. The dimension of transformer architecture is 256 and number of queries is 100. The **SFG** module takes in all 432-channels hyperspectral image and generates 1-channel output map of same spatial dimension as input. The feature extractor in **SFG** is ResNet-50 [16] initialized with DETR [4] trained on COCO dataset [30]. The decoder output embeddings are of size 512. The feature pyramid network in mask prediction module has 3 layers. More details are mentioned in supplementary materials.

	Methods	Back bone	SFG F.Ext.	#params	mAP	mIOU
JPL-CH4-detection-v1.0 Dataset						
1	Hu et. al	R-50	-	75M	0.26	0.48
2	H-mrcnn	R-50	-	353M	0.53	0.86
3	MM	R-50	R-50	**80M**	**0.63**	**0.91**
MHS (Ours) Dataset						
4	SpectralFormer	R-50	-	84M	0.33	0.41
5	UPSNet (stuff)	R-50	-	69M	0.32	0.38
6	UPSNet (stuff + things)	R-50	-	69M	0.29	0.35
7	DETR	R-18	*	33M	0.37	0.56
8	DETR	R-50	*	59M	0.44	0.59
10	MM	R-18	Linear Layer	39M	0.45	0.60
11		R-18	R-18	44M	0.52	0.63
12		R-50	R-50	**80M**	**0.59**	**0.68**

Table 2. *Comparison with baselines. "-" represent Not Applicable and "*" represent no **SFG** module and a random query used for transformer decoder. The top section shows performance on JPL-CH_4 dataset [39]. MethaneMapper achieves better results than heavily tuned H-mrcnn with $\sim 5\times$ fewer parameters. The overall detection accuracy is higher on this dataset because the type of ground terrain is uniform across all flightlines. In MHS dataset, MM outperforms multiple baselines as shown in rows 4-12. MM accuracy is lower in MHS than JPL-CH_4 dataset because MHS dataset has more variety of ground terrain spreading over 6 states*

Methods	mAP	mIOU
LogReg [5]	-	0.05
SVM [36]	-	0.29
PCA + LogReg	-	0.06
PCA + SVM	-	0.31
MM (R-50)	**0.63**	**0.91**

Table 3. *Comparison with classical machine learning methods. "-" represent Not Available. The classical ML methods are not suited for the CH_4 detection task. MethaneMapper outperforms all methods on JPL dataset [39]*

6. Results

In this section we will discuss and validate all the design choices for MethaneMapper (MM) with ablations. We show that MM achieves state-of-the-art results in overall performance compared all other methods shown in Tables 2 & 3.

6.1. Performance comparison

Deep Learning methods: We trained MM with ResNet-50 [16] backbone on the same dataset that H-mrcnn [29] (JPL-CH4-detection-V1.0 [39]) was trained on for fair comparison. To align with H-mrcnn we used the same split and input image size. The MM model with 80M parameters trained for 250 epochs outperforms by significant margin the H-mrcnn model with 352M parameters. Results are summarized in Table 2 that includes the performance of MM on the new larger MHS dataset. We note that though the code for H-mrcnn is available, many of the modules are deprecated and can not be reproduced. The 'Backbone' column represents backbones used for feature extraction from input image,'SFG F.Ext.' represents the feature extractor in **SFG** module in MethaneMapper. We observed (qualitatively) that H-mrcnn fails to detect small CH_4 plumes with concentration lower than 100kg/hr while MM detects those.

We did evaluation by implementing 3 baseline models [4, 17, 45] shown rows 4-8 of Table 2. These methods were not designed for CH_4 detection task, therefore we needed to modify their input channel size. The poor performance of these methods may be attributed to the weak signal of interest in a high dimensional data, high number of confusers, and limited annotated data. Additionally, the only hyperspectral baseline method SpectralFormer [17] has low efficiency due its pixel-wise training scheme.

Classical ML methods: We trained and tested multiple existing machine learning based approaches that are used for methane detection, performance shown in Table 3. Logistic regression (LogReg) [5] and multinomial logistic regression (MLR) [23] failed to produce any meaningful detection with 90% false positive detections. We also trained a Support Vector Machine (SVM) [36,37] based classifier, it performed slightly better than LR and MLR methods with an IOU of 21%. SVMs are prone to false positives detections same as Gaussian Mixture Models [36]. We observed that all traditional methods are not suited for the task of CH_4 detection. We also tested reducing the dimension using principal component analysis (PCA) or just taking bands which shows maximum CH_4 absorption. In the later case, the traditional methods performed better than using all 432 bands, this backs our idea of just using bands from SWIR region.

Qualitative results. Fig. 5 shows comparison of MM's mask and bounding box prediction with ground truth mask on different ground terrains. The Leakages are from different type of sources such as, oil refinery, pipeline and storage tank. MM makes correct predictions in varying scenarios.

6.2. Ablation Studies

We did the experiments for ablation on MHS dataset with ResNet-50 as backbone and validate the design choices. One parameter is changed for each ablation and others kept at best settings. More ablations in Supplementary.

Spectral Feature Generator Module: In Table 2 lower section, we show the effectiveness of our **SFG** module for the query refiner block. Our baseline is standard implementation of DETR [4] for segmentation task represented by row-1 and row-2 of Tab. 2 lower section. Using CH_4 candidates feature from **SFG** improves the bounding box detection performance by 0.14 mAP and mask prediction by 0.09 mIOU. This demonstrate that guiding queries with CH_4 candidates feature generated by **SFG** produces better embeddings as compared to random queries.

Along with this, we explored the provision of CH_4 candidates feature at 2 places, (i) at input level concatenating

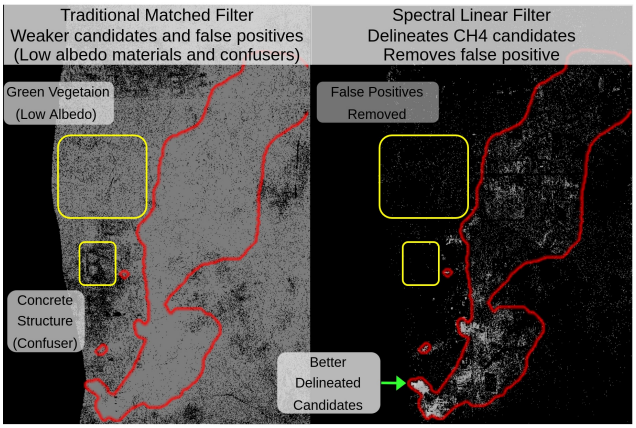

Figure 4. *Comparison of SLF with traditional filter in SFG module. White pixels represent methane and black no-methane. Red boundary represents ground-truth plume mask. SLF module generates better CH_4 candidates*

it with f_{comb}; and (ii) as input to query refiner. We see an improvement of 0.09 mAP and 0.08 mIOU when **SFG** module output is passed to query refiner. We hypothesize that this is because on concatenating with input features, the CH_4 candidates feature information gets lost, while as cross-attention with queries reduces the search space for decoder and generate better embeddings.

We also experimented with different types of feature extractors for **SFG** module, and observed that a Resnet18 or Resnet50 [16] is more effective than a 2 linear layer feature extractor as shown in Table 2.

Spectral Linear Filter: We experimented with SLF for computing covariance (*Cov*) using different subset of columns in the input hyperspectral image. We observed that the SLF is most effective when covariance is computed class-wise based on land cover. Class-wise *Cov* ensures that the radiance absorption by ground terrain is same for all the pixels while computing CH_4 enhancement. As can be seen in Fig. 4, SLF amplifies CH_4 candidate detection and reduces false positives. SLF leads to a 0.03 mAP improved in detection compared to traditional filters. The prediction from MM is shown row-1 of Fig. 5.

Geographic generalization: To assess the geographical generalization capabilities of MM, we trained it on MHS data from all states except California and tested it on flightlines from California. We observed a slight drop of 0.04 mAP in detections. However, when trained on all data except Virginia, we noticed a significant drop of 0.09 mAP in detections. We attribute this to the fact that the land cover in Virginia is dense and moist vegetation, has a lower solar reflectance compared to the arid regions of California, Texas, and Nevada.

Temporal generalization: Testing MM on 2015 after training on data from 2016-2022 showed no performance drop.

Unbalanced test set: MM's performance dropped by 0.05 mAP on an unbalanced test set with only 10% positive samples (CH_4) and 90% negative samples (no-CH_4). This highlights the challenges in CH_4 detection. Future work will address this issue.

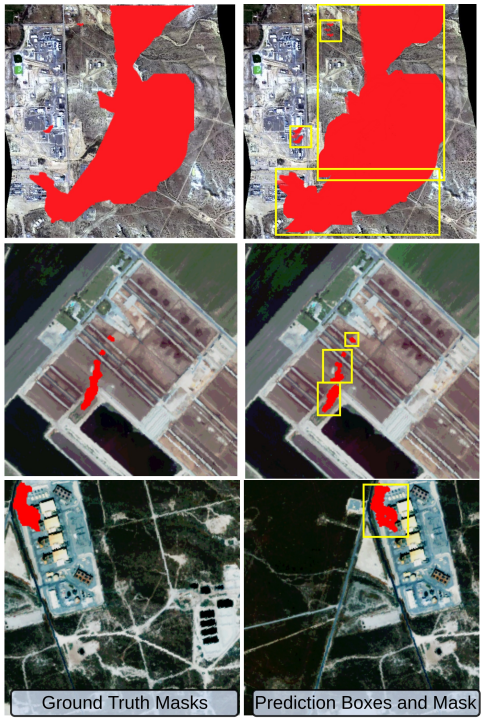

Figure 5. *Sample ground truths and predictions on MHS dataset. We show robustness of MethaneMapper predictions on different kind of ground terrain, rows 1 and 3 shows leakage at a refinery, row 2 shows leakage from pipeline in agricultural land, row 4 shows leakage from storage tank with concrete background.*

7. Conclusion

This paper presents MethaneMapper – a hyperspectral Transformer for methane plume detection. It utilize spectral and spatial correlations using a spectral feature generator and a query refiner, to accurately delineate the CH_4 plumes. Additionally, we curated a large-scale dataset for the task, a first of its kind, which will be made available to all researchers. The proposed MethaneMapper significantly improves upon the current methods in terms of detection and localization accuracy, as our extensive experiments demonstrate. Future work will extend the model to global monitoring [27] using multispectral satellite imaging data.

8. Acknowledgments

This research is partially supported by the following grants: NSF award SI2-SSI #1664172 and the US Army Research Laboratory (ARL) under agreement number W911NF2020157. We also thank Aditya Ramakrishnan in helping with the initial data cleaning process.

References

[1] Alana K. Ayasse, Andrew K. Thorpe, Dar A. Roberts, Christopher C. Funk, Philip E. Dennison, Christian Frankenberg, Andrea Steffke, and Andrew D. Aubrey. Evaluating the effects of surface properties on methane retrievals using a synthetic airborne visible/infrared imaging spectrometer next generation (aviris-ng) image. *Remote Sensing of Environment*, 215:386–397, 2018. 1

[2] CALIFORNIA AIR RESOUCE BOARD. Green house gas inventory by california air resource board, 2022. 6

[3] INC CARBON MAPPER. Carbon mapper methane emission exploration, 2022. 2, 6

[4] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020. 1, 3, 4, 5, 6, 7

[5] Qi Cheng, Pramod K Varshney, and Manoj K Arora. Logistic regression for feature selection and soft classification of remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 3(4):491–494, 2006. 7

[6] E. Knapp David, Heckler Joseph, Seely Megs, and P. Asner Gregory. Global airborne observatory visible to infrared imaging spectrometer report, 2020. 2, 6

[7] C Frankenberg, U Platt, and T Wagner. Iterative maximum a posteriori (imap)-doas for retrieval of strongly absorbing trace gases: Model studies for ch 4 and co 2 retrieval from near infrared spectra of sciamachy onboard envisat. *Atmospheric Chemistry and Physics Discussions*, 4(5):6067–6106, 2004. 2

[8] Christian Frankenberg, U Platt, and T Wagner. Iterative maximum a posteriori (imap)-doas for retrieval of strongly absorbing trace gases: Model studies for ch 4 and co 2 retrieval from near infrared spectra of sciamachy onboard envisat. *Atmospheric Chemistry and Physics*, 5(1):9–22, 2005. 1, 2

[9] Christian Frankenberg, Andrew K Thorpe, David R Thompson, Glynn Hulley, Eric Adam Kort, Nick Vance, Jakob Borchardt, Thomas Krings, Konstantin Gerilowski, Colm Sweeney, et al. Airborne methane remote measurements reveal heavy-tail flux distribution in four corners region. *Proceedings of the national academy of sciences*, 113(35):9734–9739, 2016. 1, 2

[10] Christopher C Funk, James Theiler, Dar A Roberts, and Christoph C Borel. Clustering to improve matched filter detection of weak gas plumes in hyperspectral thermal imagery. *IEEE transactions on geoscience and remote sensing*, 39(7):1410–1420, 2001. 2

[11] Bo-Cai Gao. Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, 58(3):257–266, 1996. 4

[12] Utsav B Gewali, Sildomar T Monteiro, and Eli Saber. Machine learning based hyperspectral image analysis: a survey. *arXiv preprint arXiv:1802.08701*, 2018. 2

[13] IE Gordon, LS Rothman, RJ Hargreaves, R Hashemi, EV Karlovets, FM Skinner, EK Conway, C Hill, RV Kochanov, Y Tan, et al. The hitran2020 molecular spectroscopic database. *Journal of quantitative spectroscopy and radiative transfer*, 277:107949, 2022. 4

[14] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 4

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015. 6

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3, 4, 6, 7, 8

[17] Danfeng Hong, Zhu Han, Jing Yao, Lianru Gao, Bing Zhang, Antonio Plaza, and Jocelyn Chanussot. Spectralformer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2021. 2, 7

[18] Paris IEA. Methane from oil & gas, 2020. 1

[19] ASM Iftekhar, Hao Chen, Kaustav Kundu, Xinyu Li, Joseph Tighe, and Davide Modolo. What to look at and where: Semantic and spatial refined transformer for detecting human-object interactions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5353–5363, 2022. 3, 4

[20] ASM Iftekhar, Satish Kumar, R Austin McEver, Suya You, and BS Manjunath. Gtnet: Guided transformer network for detecting human-object interactions. *arXiv preprint arXiv:2108.00596*, 2021. 3, 4

[21] California Institute of Technology Jet Propulsion Laboratory. Airborne visible infrared imaging spectrometer - next generation (aviris-ng) overview, 2009. 1, 2, 5

[22] Siraput Jongaramrungruang, Andrew K Thorpe, Georgios Matheou, and Christian Frankenberg. Methanet–an ai-driven approach to quantifying methane point-source emission from high-resolution 2-d plume imagery. *Remote Sensing of Environment*, 269:112809, 2022. 1, 2

[23] Mahdi Khodadadzadeh, Jun Li, Antonio Plaza, and José M Bioucas-Dias. A subspace-based multinomial logistic regression for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 11(12):2105–2109, 2014. 7

[24] Stefanie Kirschke, Philippe Bousquet, Philippe Ciais, Marielle Saunois, Josep G Canadell, Edward J Dlugokencky, Peter Bergamaschi, Daniel Bergmann, Donald R Blake, Lori Bruhwiler, et al. Three decades of global methane sources and sinks. *Nature geoscience*, 6(10):813–823, 2013. 1

[25] Satish Kumar, ASM Iftekhar, Michael Goebel, Tom Bullock, Mary H MacLean, Michael B Miller, Tyler Santander, Barry Giesbrecht, Scott T Grafton, and BS Manjunath. Stressnet: detecting stress in thermal videos. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 999–1009, 2021. 3

[26] Satish Kumar, ASM Iftekhar, Ekta Prashnani, and BS Manjunath. Locl: Learning object-attribute composition using localization. *arXiv preprint arXiv:2210.03780*, 2022. 3

[27] Satish Kumar, William Kingwill, Orbio Earth, Rozanne Mouton, Wojciech Adamczyk, Robert Huppertz, and Evan Sherwin. Guided transformer network for detecting methane emissions in sentinel-2 satellite imagery. 8

[28] Satish Kumar, Rui Kou, Henry Hill, Jake Lempges, Eric Qian, and Vikram Jayaram. In-situ water quality monitoring in oil and gas operations. *arXiv preprint arXiv:2301.08800*, 2023. 2

[29] Satish Kumar, Carlos Torres, Oytun Ulutan, Alana Ayasse, Dar Roberts, and BS Manjunath. Deep remote sensing methods for methane detection in overhead hyperspectral imagery. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1776–1785, 2020. 1, 2, 6, 7

[30] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 6

[31] META-STANFORD. Methane emissions technology alliance (meta), stanford natural gas initiative, 2022. 2

[32] MGISGeography. Ndvi (normalized difference vegetation index), 1979. 4

[33] Gunnar Myhre, Drew Shindell, and Julia Pongratz. Anthropogenic and natural radiative forcing. 2014. 1

[34] Darius Phiri, Matamyo Simwanda, Serajis Salekin, Vincent R Nyirenda, Yuji Murayama, and Manjula Ranagalage. Sentinel-2 data for land cover/use mapping: a review. *Remote Sensing*, 12(14):2291, 2020. 2, 6

[35] Dar A Roberts, Eliza S Bradley, Ross Cheung, Ira Leifer, Philip E Dennison, and Jack S Margolis. Mapping methane emissions from a marine geological seep source using imaging spectrometry. *Remote Sensing of Environment*, 114(3):592–606, 2010. 1, 2

[36] CA Shah, PK Varshney, and MK Arora. Ica mixture model algorithm for unsupervised classification of remote sensing imagery. *International Journal of Remote Sensing*, 28(8):1711–1731, 2007. 2, 7

[37] David MJ Tax and Robert PW Duin. Support vector data description. *Machine learning*, 54(1):45–66, 2004. 2, 7

[38] DR Thompson, I Leifer, H Bovensmann, M Eastwood, M Fladeland, C Frankenberg, K Gerilowski, RO Green, S Kratwurst, T Krings, et al. Real-time remote detection and measurement for airborne imaging spectroscopy: a case study with methane. *Atmospheric Measurement Techniques*, 8(10):4383–4397, 2015. 1, 2

[39] David R Thompson, Anuj Karpatne, Imme Ebert-Uphoff, Christian Frankenberg, Andrew K Thorpe, Brian D Bue, and Robert O Green. Isgeo dataset jpl-ch4-detection-2017-v1. 0: A benchmark for methane source detection from imaging spectrometer data. 2017. 2, 5, 6, 7

[40] Andrew K Thorpe, Christian Frankenberg, David R Thompson, Riley M Duren, Andrew D Aubrey, Brian D Bue, Robert O Green, Konstantin Gerilowski, Thomas Krings, Jakob Borchardt, et al. Airborne doas retrievals of methane, carbon dioxide, and water vapor concentrations at high spatial resolution: application to aviris-ng. *Atmospheric Measurement Techniques*, 10(10):3833–3850, 2017. 1, 2, 4

[41] Andrew K Thorpe, Dar A Roberts, Eliza S Bradley, Christopher C Funk, Philip E Dennison, and Ira Leifer. High resolution mapping of methane emissions from marine and terrestrial sources using a cluster-tuned matched filter technique and imaging spectrometry. *Remote Sensing of Environment*, 134:305–318, 2013. 1, 2, 4

[42] George Turin. An introduction to matched filters. *IRE transactions on Information theory*, 6(3):311–329, 1960. 4

[43] Oytun Ulutan, ASM Iftekhar, and Bangalore S Manjunath. Vsgnet: Spatial attention network for detecting human object interactions using graph convolutions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13617–13626, 2020. 3

[44] Foreign Agriculture Service US Dept. of Agriculture. Normalized difference vegetation index (ndvi), 1969. 4

[45] Yuwen Xiong, Renjie Liao, Hengshuang Zhao, Rui Hu, Min Bai, Ersin Yumer, and Raquel Urtasun. Upsnet: A unified panoptic segmentation network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8818–8826, 2019. 7