Supplementary Materials
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

1. Introduction

In supplementary section, we provide will all the details about the data collection and annotations creation process. We also provide with the complete derivation of Spectral Linear Filter (**SLF**) along with a pseudo implementation of **SLF** algorithm. Next in the document we provide some more qualitative examples of success and failure cases of MethaneMapper. Towards the end of the document we provide graph plots about training convergence of all the ablation experiments with Spectral Feature Generator (SFG) and Query Refiner (QR) module.

2. Dataset

2.1. AVIRIG-NG

AVIRIS-NG [5] is an acronym for the *Airborne Visible InfraRed Imaging Spectrometer - Next Generation* developed by Jet Propulsion Laboratory (JPL) in 2009. JPL conducted thousands of flight lines recording data with AVIRIS-NG instrument in last 7 years. On the AVIRIS-NG instrument an array of total 598 sensors in push-broom order captures an unortho-rectified data-cube of spatial dimension $\sim 23k \times 598$, where each sensor records a spectral wavelengths ranging from $380nm - 2510nm$ [4] making a dimension of 432 channels. It has 34^o field of view with a 1 mrad instantaneous field of view the generates spatial resolution of $1-8m$ based on altitude. This data is then rectified using a geometric lookup table and the resulting data cube is of size $\sim 23k\times \sim 1.5k \times 432$. The data is provided in Band Interleaved by Line (BIL) ordering. BIL ordering signifies the 3D matrix is indexed first by image row, then by channel, and then by the image column [12]. One can find details about the naming convention and the type of data each files contain in "README.txt" file in each flightline folder. The data can be loaded into a *numpy* array easily using python libraries. All data is orthorectified.

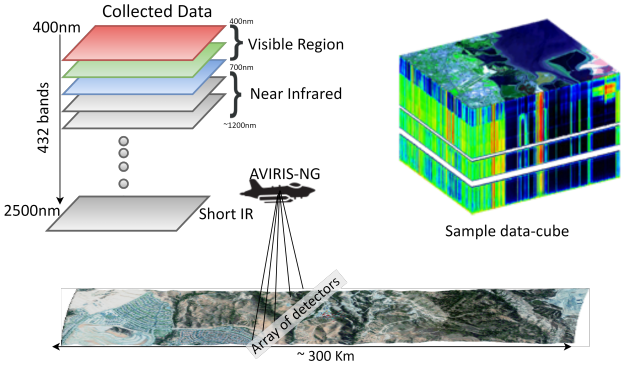

Figure 1. *Depiction of data collection process. Each flightline is* $\sim$ 300 *kms long. An array of 598 sensors records data at 1.5m/pixel spatial resolution. All flightlines are ortho-corrected. Each data-cube is of dimension* $\sim 23k\times \sim 1.5k \times 432$.

2.2. Annotations

Transformation and Ortho-correction. First step is to read the annotation GeoTiff patch of size 150×150 of a methane concentration mask and convert its Coordinate Reference System (CRT) to AVIRIS-NG flightlines' CRT (EPSG 4326). Next, we use the corresponding AVIRIS-NG flightlines' geometric lookup table and unortho-corrected geographic pixel location to generate ortho-corrected geographic pixel location data of the flightline. Next, we find the flightline's geographic indices that are closest to the geographic indexes of the methane concentration mask (annotation GeoTiff). Finally, we use these corresponding pixels to compute a homography transform matrix that maps the methane concentration mask (annotation GeoTiff) to the AVIRIS-NG flightline's spatial dimensions. We repeat this process for each plume in the flightline in order to generate the CH_4 concentration map for the entire flightline.

Resolution matching. To match the resolution of transformed annotation GeoTiff patch to AVIRIS-NG flightline,

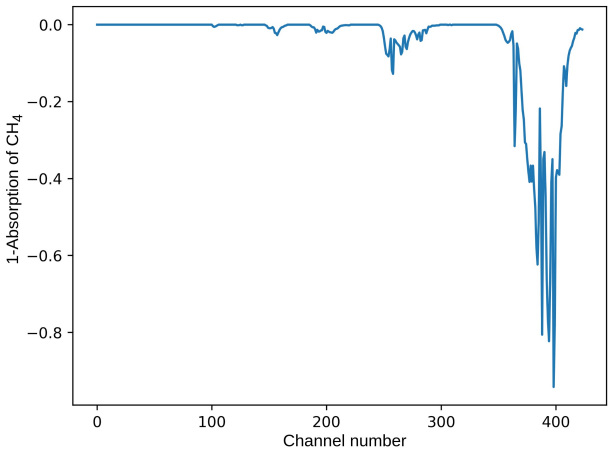

Figure 2. *Spectral absorption pattern of CH_4 gas. The x-axis show the channel number ranging from 0-400 corresponding to wavelength range ($400nm - 2500nm$). It is obtained from the public repository HITRAN [3].*

we use nearest-neighbor resampling. A pixel from the transformed annotation GeoTiff patch may be repeated multiple times in the CH_4 concentration map for the entire flightline.

Annotation Style. The *Point Source* and *Diffused Source* are coded following the same standard as JPL-CH4-detection-V1.0 [12] dataset. The 3-channels have values in [0-255] range.

- Red (255,0,0): plume, believed to be associated with a *Point Source*
- Blue (0,0,255): plume, believed to be associated with a *Diffuse Source*
- Black (0,0,0): no plume (or unlabeled)

We kept our annotation style consistent with JPL-CH4-detection-V1.0 benchmark dataset [12] so that both JPL-CH4-detection-V1.0 and MHS datasets can be merged seamlessly.

3. Spectral Linear Filter(SFL)

3.1. Traditional Matched Filter

Passive hyperspectral imaging sensors captures spectral radiances values from N_0 ($N_0 = 432$) spectral channels corresponding to wavelengths ranging from $400nm - 2500nm$ as shown in Fig. 1 with sample data-cube. The complete hyperspectral image is represented as $\mathbf{x} \in \mathbb{R}^{H_0 \times W_0 \times N_0}$ where $H_0, W_0 \& N_0$ are height, width and number of channels respectively. In this hyperspectral data, we are looking for a very weak signature of interest hidden in background. In this case the signature of interest is CH_4 and the background is ground terrain. CH_4 shows strong absorption patterns around $2100nm - 2500nm$ wavelength.

The most common linear approach for finding CH_4 candidates is taking a N_0-dimension (same as number of spectral channels) vector α, and apply as a dot product to each pixel (N_0-dimension) in the hyperspectral image to generate a scalar output per pixel. This operation is supposed to reduce or remove the ground terrain, sensor noise and amplifies CH_4 signature. The α vector used here is called as "matched filter". Therefore computing right α is very critical for generating better candidates of CH_4 emission. It is dependent on absorption pattern of CH_4 and on the distribution of the ground terrain. To model α, let $\mathbf{r}_i \in \mathbb{R}^n$ be a i^{th} pixel from the hyperspectral image representing the ground terrain pixel and sensor noise, and $\mathbf{t}$ be the CH_4 absorption pattern [3]. This is modeled as the additive perturbation as shown below:

$$\mathbf{x}_i = \mathbf{r}_i + \mathbf{t}, \tag{1}$$

where $\mathbf{x}_i$ is the spectrum when CH_4 is present. The CH_4 absorption pattern $\mathbf{t}$ represents the change in radiance units of the background caused by adding a unit mixing ratio length of CH_4 absorption [1, 7]. Figure 2 shows the spectral absorption pattern of CH_4 per channel. In the ideal scenario where only CH_4 gas is present in signal (i.e. all white background), the matched filter output is $\alpha^T\mathbf{t}$. In case there is no gas and just ground terrain and sensor noise, the matched filter output is $\alpha^T\mathbf{r}_i$. The variance (Var) of $\alpha^T\mathbf{r}_i$ for latter is represented by:

$$Var(\alpha^T\mathbf{r}_i) = \langle(\alpha^T\mathbf{r}_i - \alpha^T\boldsymbol{\mu})^2\rangle = \alpha^T\mathbf{Cov}\alpha, \tag{2}$$

where $\mathbf{Cov}$ and μ are covariance and mean respectively computed for $\mathbf{r}_i$. Inspired from [1, 7] we define the Methane-to-Ground terrain Ratio (MGR) is:

$$\text{MGR} = \frac{|\alpha^T\mathbf{t}|^2}{\alpha^T\mathbf{Cov}\alpha}, \tag{3}$$

We can see that the magnitude of α does not affect MGR. According to [1, 7, 11], the MGR can be maximized subject to constraints(zero mean and $\alpha^T\mathbf{K}\alpha$ constraint to 1). The matched filter α is then represented by:

$$\alpha = \frac{\mathbf{Cov}^{-1}\mathbf{t}}{\sqrt{\mathbf{t}^T\mathbf{Cov}^{-1}\mathbf{t}}}. \tag{4}$$

In ideal instances when there is no background (i.e. all white background) and just CH_4 gas present. The matched filter in equation 5 is directly proportional to $\mathbf{t}$. This is just the target signature ($\mathbf{t}$) itself scaled so that the filtered output has variance of one. The methane enhancement per pixel can be computed as follows:

$$\hat{\alpha}(\mathbf{x}_i) = \frac{(\mathbf{x}_i - \mu)^T\mathbf{Cov}^{-1}\mathbf{t}}{\sqrt{\mathbf{t}^T\mathbf{Cov}^{-1}\mathbf{t}}}, \tag{5}$$

where $\hat{\alpha}(\mathbf{x}_i)$ is the per pixel estimation of methane, on other words, column enhancement of methane. The covariance

matrix (**Cov**) used is not known as $prior$ and is estimated from data. It is computed as outer product of the mean subtracted radiance over all the pixels. In other words, the traditional matched filter from equation 5 computes the covariance (**Cov**) of ground terrain with an underlying assumption that in all elements have similar absorption pattern. Same covariance matrix (**Cov**) matrix is used to whitens the varying ground terrain and amplify the CH_4 present. But in realistic scenarios, the ground terrain is varying, the type of terrain changes frequently, there is water bodies, bare soil, vegetation, dense vegetation, building structures in cities, roads etc in a single image. For example, water have a strong absorption of solar radiations, therefore the methane on such backgrounds have a very weak visibility. Similarly, wet fields dense vegetation have similar behaviour. On the other hand, bare soil, rocks, etc have lower absorption, the methane present on such background have strong visibility. A simple and single approximation of the covariance (**Cov**) of ground distribution can not provide the right and effective estimate of methane enhancement. To tackle this limitation, we developed an spectral linear Filter (SLF) that does land cover classification and segmentation and reduces the noise as discussed in the next sections.

3.2. Landcover Classification and Segmentation

In this section, we improve upon the limitations mentioned in the previous section. We start with taking hyperspectral bands from visible spectrum ($400nm-700nm$) and near-mid infrared region ($800nm-1350nm$). We recreated the RGB representation of the ground terrain by a weighted normal distribution for each color band. Same is done for near infrared region. Next we take a simple, very effective and efficient approach for doing landcover classification and segmentation. We compute the Normalized Difference Vegetation Index (NDVI) [9, 10] and Normalized Difference Water Index (NDWI) [2]. NDVI quantifies vegetation by measuring the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs) [9]. It ranges from -1 to $+1$. It is a very effective index and has been used in literature for more than 4 decades. [2] created NDWI and used it to highlight open water features in a satellite image, allowing a water body to "stand out" against the soil and vegetation. It is calculated using the GREEN-NIR (visible green and near-infrared) and ranges from -1 to $+1$. Its primary use today is to detect and monitor slight changes in water content of the water bodies.

$$ndvi = \frac{NIR - R}{NIR + R}; \quad ndwi = \frac{NIR - MIR}{NIR + MIR} \tag{6}$$

where NIR is near infrared region normalized around $880nm$, MIR is mid infrared normalized around $1240nm$ and R is red, normalized around $660nm$. We take advantage of these indexes and create segmentation maps for different types of vegetation, water bodies, bare soil, rocks, mountains, city/urban areas, roads etc. We take the classification thresholds from [6, 13]. For simplification, we also tested by splitting the scale -1 to $+1$ in 20 classes, each with a range of $< 0.1 >$. We obtained comparable results as compared to using classification ranges from [6,13]. This simple, effective and efficient approach gives three fold boost to our spectral linear filter CH_4 candidates estimation.

3.3. Cov per class

We take the segmented image from previous step, we will call segmented image as segmentation mask for simplicity now onward. In practice we have 20 classes, each with a segmentation mask. We merged two or more adjacent classes into one if the number of pixels in that class is less 10000 . The Number of pixels in each class is kept higher to ensure that while computing the covariance (**Cov**) matrix, the methane signal does not have any or have negligible effect. It is okay to merge adjacent classes into one because they have almost similar radiance/reflectance, for example, light vegetation and normal vegetation have similar reflectance, etc. For each class we compute a separate mean and covariance matrix. The covariance $\mathbf{Cov}_k$ of k^{th} class is computed as:

$$\mathbf{Cov}_k = \frac{1}{N}\sum_{i=1}^{i=j}(\mathbf{x}_i - \mu_k)(\mathbf{x}_i - \mu_k)^T \; \forall \; j \in k, \tag{7}$$

where N is the number of pixels (> 10000) in k^{th} class and μ_k is the mean of k^{th} class. For each class we compute the mean μ_k, covariance matrix $\mathbf{Cov}_k$. While iterating through each pixel of hyperspectral image, we check to which class k the pixel $\mathbf{x}_i$ belongs to and use those pre-computed values. The final Spectral Linear Fitler (**SLF**) is shown as below:

$$\mathbf{SLF}(\mathbf{x}_i) = \frac{(\mathbf{x}_i - \mu_k)^T \mathbf{Cov}_k^{-1} t}{\sqrt{t^T \mathbf{Cov}_k^{-1} t}} \; \forall \; (i) \in \text{class } k \tag{8}$$

where $\mathbf{Cov}^{-1}$ is the inverse of covariance matrix. Next to suppress the sensor noise, we exploit the simple method of tracking each sensor. Each sensor have different physical properties, that can influence the data captured by it. We track each individual sensor in the flight line. Since the data is rectified, the data from each sensor does not belong to single column, instead it is spread randomly across all the columns. This is dependent on the flight path and the movement in the airplane while moving. We used simple data-structure algorithms like depth first search. Tracked each boundary pixels and assigned them to single sensor. We used data from 10-15 adjacent sensor at one time, normalize it and then compute the covariance matrix in pre-

vious step with segmentation mask. Our approach is very simple and straight forward.

The algorithm 1 shows the pseudo code for our Spectral Linear Filter (**SLF**).

Data: *MHS dataset*
Result: *CH_4 candidates*
initialization;
for mhs *in* MHS **do**
  1. create memory map mhs;
  2. seg_mask = compute segmentation mask;
  for *mask in seg_mask* **do**
    data.**append**(mhs[mask])
    if (len(data) $<$ 100000): **continue**
    Cov, μ = compute_stats(data);
  end
  3. sensor_array = individual sensors;
  for *arrays in sensor_array* **do**
    data = mhs[arrays]
    for x_i *in data* **do**
      k = seg_mask[i];
      SLF $(\mathbf{x}_i) = \frac{(\mathbf{x}_i - \mu_k)^T \mathbf{Cov}_k^{-1} t}{\sqrt{t^T \mathbf{Cov}_k^{-1} t}}$ **end**
    end
    SLF $(\mathbf{x}_i)$ $\forall$ $classes$ and $i \in mhs$
  end

Algorithm 1: Spectral Linear Filter (**SLF**)

3.4. Training policy

We trained MethaneMapper in two styles, (i) pre-training the bounding box and class detection first and then freezing the pre-trained model parameters and training only the mask prediction layer; and (ii) trained whole pipeline end-to-end and achieved similar performance on both the cases.

3.5. Qualitative Results

In this section we show few more qualitative examples of CH_4 plume mask prediction and few cases where MM failed to detect any CH_4 gas emission.Figure. 3 shows the CH_4 detections in different types of background terrain and different types of emission source.

Figure 4 shows some examples of missed CH_4 plume detections. We observed that going back to dataset samples and checking the timelines, these flightlines were recorded during the evening time. We believe that this might be because of evening time, the reflectance from the ground terrain is very weak and small. Hence we believe there is minimum absorption of reflected solar radiation by CH_4 gas present in the atmosphere and the plume goes undetected.

4. Ablations Studies

Attention Type: We also explored different attention mechanisms to encode and decode information. We replaced only the attention layers with deformable-attention [14] in the our architecture that resulted in a drop of 0.1 mAP in the baseline model.

5. Implementation details

The whole network is trained with AdamW [8] optimizer, batch size of 12, with initial learning rate for backbones set to 10^{-5} and for transformer the learning rate is set to 10^{-5} with a weight decay of 10^{-4}. The learning rate for mask prediction module is set to 10^{-4}. The learning rate is dropped at every 150 epochs, we train for 300 epochs. The baseline model is trained on 2 V100 GPUs.

6. 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.

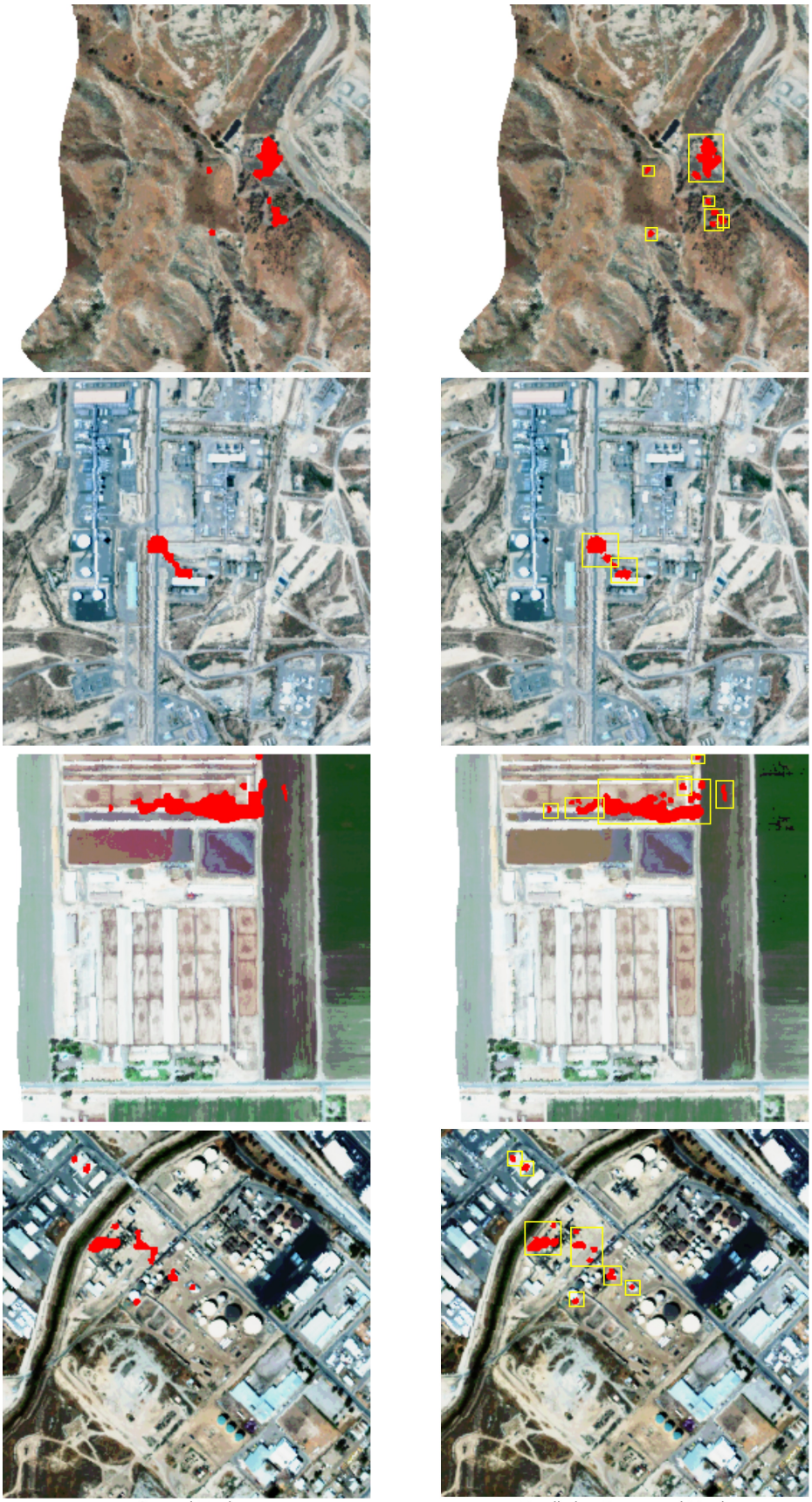

Figure 3. *Sample ground truths and predictions on MHS dataset. We are showing different type of terrains and CH_4 predictions on them. The type of emission source in all samples varies too.*

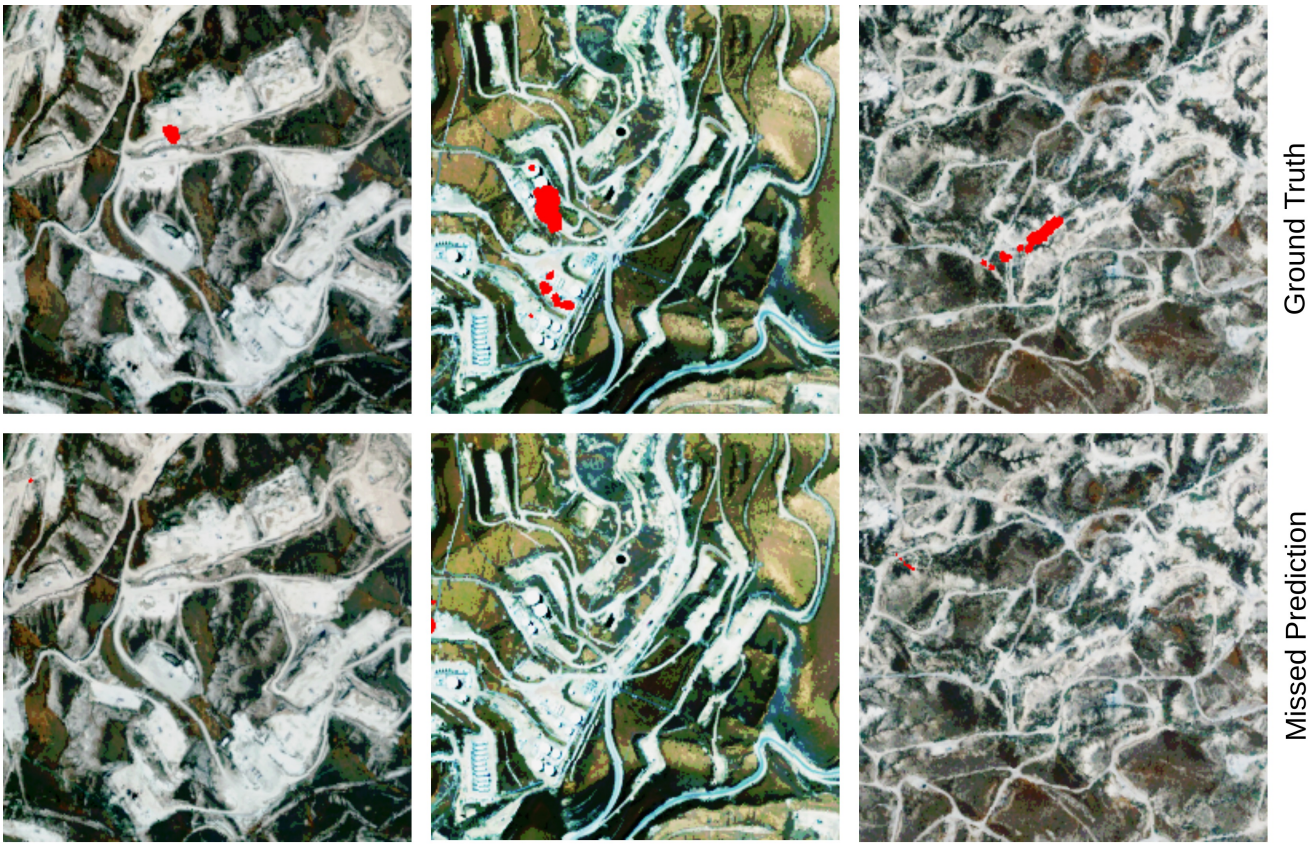

Figure 4. *Samples where MM fails to detect the CH_4 plume. We observed that these samples were recorded during the evening time and hence reflectance from the ground terrain is very weak. Therefore the absorption of reflected solar radiations by CH_4 is very low and hence the emissions goes undetected.*

References

[1] 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

[2] 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. 3

[3] 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. 2

[4] L. Hamlin, R. O. Green, P. Mouroulis, M. Eastwood, D. Wilson, M. Dudik, and C. Paine. Imaging spectrometer science measurements for terrestrial ecology: Aviris and new developments. In *2011 Aerospace Conference*, pages 1–7, 2011. 1

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

[6] FJ Kriegler, WA Malila, RF Nalepka, and W Richardson. Preprocessing transformations and their effects on multispectral recognition. *Remote sensing of environment, VI*, page 97, 1969. 3

[7] 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. 2

[8] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 4

[9] MGISGeography. Ndvi (normalized difference vegetation index), 1979. 3

[10] Nathalie Pettorelli. *The normalized difference vegetation index*. Oxford University Press, 2013. 3

[11] James Theiler, Bernard R Foy, and Andrew M Fraser. Beyond the adaptive matched filter: nonlinear detectors for weak signals in high-dimensional clutter. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII*, volume 6565, pages 26–37. SPIE, 2007. 2

[12] 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. 1, 2

[13] Foreign Agriculture Service US Dept. of Agriculture. Normalized difference vegetation index (ndvi), 1969. 3

[14] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*, 2020. 4