Deep Remote Sensing Methods for Methane Detection in Overhead Hyperspectral Imagery

Satish Kumar[†*] Carlos Torres[†*] Oytun Ulutan[†] Alana Ayasse[‡] Dar Roberts[‡] B.S. Manjunath[†]
University of California Santa Barbara [†]ECE Department [‡]Geography Department
{satishkumar@, carlostorres@ece, ulutan@, alanaayasse@, dar@geog, manj@}.ucsb.edu

Abstract

Effective analysis of hyperspectral imagery is essential for gathering fast and actionable information of large areas affected by atmospheric and green house gases. Existing methods, which process hyperspectral data to detect amorphous gases such as CH_4 require manual inspection from domain experts and annotation of massive datasets. These methods do not scale well and are prone to human errors due to the plumes' small pixel-footprint signature. The proposed Hyperspectral Mask-RCNN (H-mrcnn) uses principled statistics, signal processing, and deep neural networks to address these limitations. H-mrcnn introduces fast algorithms to analyze large-area hyper-spectral information and methods to autonomously represent and detect CH_4 plumes. H-mrcnn processes information by match-filtering sliding windows of hyperspectral data across the spectral bands. This process produces information-rich features that are both effective plume representations and gas concentration analogs. The optimized matched-filtering stage processes spectral data, which is spatially sampled to train an ensemble of gas detectors. The ensemble outputs are fused to estimate a natural and accurate plume mask. Thorough evaluation demonstrates that H-mrcnn matches the manual and experience-dependent annotation process of experts by 85% (IOU). H-mrcnn scales to larger datasets, reduces the manual data processing and labeling time ($\times 12$), and produces rapid actionable information about gas plumes.

1. Introduction

The presence of methane gas (CH_4) in the atmosphere is understood to be a chief contributor to global climate change. CH_4 is a greenhouse gas with a Global Warming Potential (GWP) 86 times that of carbon dioxide (CO_2) [25]. CH_4 accounts for 20% of global warming induced by greenhouse gases [19]. Although CH4 has many sources, oil and natural gas are of particular interest. Emissions from this sector tend to emanate from specific locations, like natural gas storage tank leaks or pipelines leaks.

[0]Satish and Carlos co-authored this paper as equal contributors.

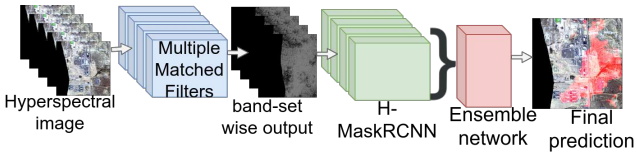


Figure 1: Overview of proposed method. From left-to-right: the hyperspectral image input (left), in sets of bands, are processed by multiple matched filters (blue block). A bank of detectors is trained on outputs from multiple matched filters (green block). The bank outputs are fused by 2-layer perceptron (red block) to give the final prediction of the plume red-overlay (right most).

These emissions exhibit plume-like morphology, which makes distinguishing them from the background both feasible and challenging. The Jet Propulsion Laboratory (JPL) collected data using the Airborne Visible/Infrared Imaging Spectrometer Next Generation (AVIRIS-NG) [13] to monitor and investigate such emissions. AVIRIS-NG captures information at wavelengths ranging from the visible spectrum to short-infrared spectrum (i.e., $380nm - 2510nm$). Information about CH_4 is present as a small signal in the $2100nm$ to $2400nm$ range. Recent work produced algorithms that detect the CH_4 signal in the AVIRIS-NG images [22, 28]. However, the outputs from these detection algorithms can be noisy and have spurious signals. Extensive manual labor is still required to identify and delineate the methane plumes. This work proposes a hybrid technique that combines core concepts of conventional signal processing and machine learning with deep learning. This technique addresses the limitations of existing methods, such as computational complexity, speed, and manual processing bottlenecks by harnessing the spatial information (i.e., plume shape) and spectral information to automatically detect and delineate CH_4 plumes in overhead imagery.

Aerial imagery is commonly used to identify sources of CH_4 i.e., point source region and estimate CH_4 concentration in large areas [9, 10, 28, 30, 31]. Remote sensing instruments such as AVIRIS-NG have high spectral resolution and are capable of detecting point sources of CH_4.

Methane Detection. Retrieval of CH_4 emission sources from hyperspectral imagery is a recent topic of study in re-

978-1-7281-6553-0/20/$31.00 ©2020 IEEE

mote sensing [9, 10, 28, 30, 31]. Hyperspectral sensors, like AVIRIS-NG, are not originally designed for gas detection but are effective tools to observe gases due to their spectral range. There are two existing methods to estimate column-wise concentration of methane from AVIRIS-NG data. The IMAP-DOAS algorithm [9] was adapted for AVIRIS-NG [30]. This method uses Beer-Lambert law, where Differential Optical Absorption Spectroscopy (DOAS) describes the relationship between incident intensity for vertical column and measured intensity after passing through a light path containing an absorber [17]. CH_4 retrievals are performed between $2215nm$ and $2410nm$. Other methods involve matched filter approaches to estimate column concentration of CH_4 [10, 28, 31, 30]. The matched filter tests the null hypothesis H_0 (spectrum generated by the background) against the alternative hypothesis H_1 (spectrum including the perturbation due to gas). The Cluster-Tuned Matched Filter algorithm [31] is used to detect the presence of CH_4 or strength of presence of CH_4 signal. This method is commonly applied to data acquired by AVIRIS-NG but the results are noisy and prone to false positives.

Related Technical Work. Existing machine learning-based hyperspectral image analysis methods mostly focus on classification with a small portion dedicated to target detection as reported in [11]. For instance, logistic regression is commonly used for land cover classification in remote sensing application using pixel-wise classification [8]. However, this method is prone to false positives. The multinomial logistic regression (MLR) [18], is a discriminative approach that directly models posterior class distributions. This type of classifier is specifically designed for the linear spectral unmixing process applications. Support vector machines (SVMs) are the most used algorithms for hyperspectral data analysis [4]. SVM generates a decision boundary with the maximum margin of separation between the data samples belonging to different classes. Target detection have been performed using a SVM related algorithm called support vector data description [26, 23]. This method generates a minimum enclosing hypersphere containing the targets. The main limitation of the method is that it does not account any underlying distribution of the scene data and fails to distinguish target from underlying background distribution. Gaussian mixture models (GMMs) represent the probability density of the data with a weighted summation of a finite number of Gaussian densities with different means and standard deviations. GMMs cluster the hyperspectral data with connected component analysis to segment the image into homogeneous areas [24].

Latent linear models find a latent representation of the data by performing a linear transform. The most common in hyperspectral imagery is PCA (Principle Component Analysis). The PCA linearly projects the data onto an orthogonal set of axes such that the projections onto each axis are uncorrelated. It is widely used as a preprocessing tool for hyperspectral analysis [7, 21, 32]. Ensemble Learning is a supervised learning technique of merging the results from multiple base predictors to produce a more accurate result. It is applied successfully for hyperspectral classification [16]. Kernelized PCA to reduce dimension followed by deep learning methods is a potential solution to target detection [6, 33, 34]. The authors from [5] introduce a three dimensional end-to-end convolutional neural network (CNN) to predict material class from the image patches (i.e., tile) around the test pixel. Three dimensional CNN outperforms the two dimensional CNN by directly learning the spatial-spectral features as their filters span over both spatial and spectral axes; however, it requires large training datasets.

Proposed H-mrcnn solution. The single-band CH_4 array is combined with the ground terrain information to train a Deep Neural Network (DNN) based detector. The naive DNN detector leverages the standard Mask-RCNN (Region Convolution Neural Network) [14] to produce a binary segmentation mask of CH_4 plumes. Mask-RCNN is suited for this problem as it looks for specific patterns in the underlying distribution. The naive DNN detector method is the basis of H-mrcnn. The raw data (432 bands) are processed in sets of bands, where H-mrcnn generates a segmentation mask (plume) for each set of bands. This ensemble of detectors (H-mrcnn) captures different distribution information from each set of bands. The detectors with input from visible and near infrared wavelength range capture the distribution of the underlying terrain. These detectors learn to eliminate the potential confusers (same signature as methane) such as hydrocarbon paints on large warehouses or asphalt roads. The detectors trained on bands with wavelength in short infrared regions capture the distribution of the CH_4 signature. The output mask candidates from the detectors are fused by a simple 2-layer perceptron network to decide a weight for each mask and its overlay. Learning methane signatures, confuser signatures, and plume and confusers shapes helps to simultaneously predict reliable plume shapes and eliminate the false positives.

Experimental results indicate that the ensemble and fusion methods are effective representations and detectors of CH_4 plumes and their shapes. The decision mechanism weights the contribution of each weak detector and produces an estimate of the gas presence or absence (overlapping the CH_4 detections from each detector in the ensemble). Thorough literature search indicates that H-mrcnn is the first solution that addresses the large-area hyperspectral data analysis problem. It introduces new methods to delineate and detect amorphous gas plumes using principled statistics, signal processing, and deep neural networks.

Technical Contributions and Innovations

1. A novel approach optimized for binary plume detection via ensembles, which better describe gas shapes.
2. Large-area data inspection and visualization tools.
3. A new and improved method to effectively use all the 432-bands (hyperspectral) data for rapid processing and analysis of hyperspectral information.
4. An autonomous plume detector that estimates binary plume masks (methane plume vs. no methane plume) and a ensemble method that estimates relative plume enhancements representations (i.e., analog) using higher resolution per-band window information.
5. An effective template for an end-to-end method to analyze and process "RAW" hyperspectral data for a specific gas signature. This study uses CH_4 as an example that is generalizable to other gas signatures.

The proposed H-mrcnn is a combination of an optimized matched filter and Mask-RCNN that identifies the correlation both in spectral and spatial domains respectively and detects the presence and shape CH_4 plume.

Extensive experimentation shows the performance of H-mrcnn compared to traditional machine learning algorithms such as logistic regression, SVM, meanshift with watershed and linear latent models and state of the art deep learning based segmentation model Learn to Segment Everything [15]. The proposed solution outperforms the competing methods in terms of detection accuracy and/or speed.

Collection of CH_4 Data. Although the AVIRIS [12] and AVIRIS-NG [13] sensors are not designed for detecting CH_4 emissions, and are used for high resolution mapping of natural CH_4 seeps [22] and fugitive CH_4 emissions [28, 31, 2]. The quantification of gas presence in a certain location is based on its atomic and molecular properties. The gases absorb a certain wavelength of light (an absorption line spectrum). CH_4 gas absorbs light in the wavelength range $2200nm$ to $2400nm$. The detection of CH_4 signal strength is based on its detected absorption (i.e., more methane yields a stronger signature).

Major challenges of plume representation and detection is their rarity and the small-pixel footprint compared to the large observed area. The occurrence frequency of the plumes in this dataset relative to the image dimensions is shown in Figure 2. This histogram shows that the highest ratio in this dataset is only 1.12 percent. The most common image-plume portion is less than 0.28% found in 36 images.

The proposed methods are developed and tested on two datasets derived from AVIRIS-NG instrument: **Dataset A** is a rectified 4-band dataset defined in [29]. The data contains 4-band datum with three bands comprising red, green, and blue reflectance intensities and a fourth band comprising

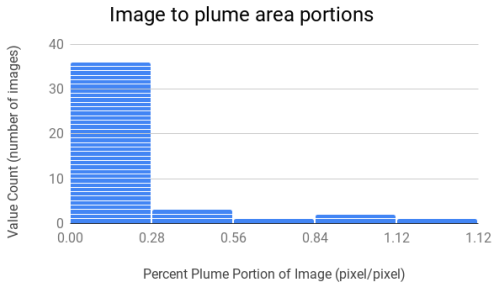


Figure 2: Frequency count plot of the percent ratio of plume-pixels to hyperspectral-pixels (rows and columns). The plot indicates that the plumes are a very small portion of the image(i.e., small pixel foot-print).

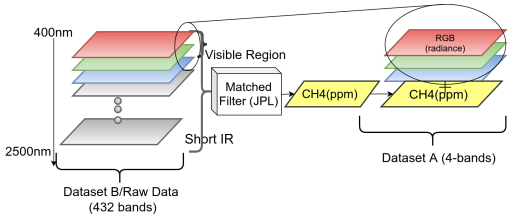


Figure 3: Relation between dataset A (χ_A) and dataset B (χ_B). The 432-bands data from dataset B are processed through a matched filter to yield dataset A. Detecting plumes using this information poor dataset (Dataset A) is challenging. H-mrcnn addresses this challenge by modeling terrain absorption using ensemble and decision fusion methods.

CH_4 relative concentration in ppm per meter (parts per million per meter); **Dataset B** is an unrectified, 432-band (i.e., raw data) dataset. It is acquired in VSWIR(Visible Shortwave Infrared) range, measuring over 432 spectra of color channels ranging from ultraviolet ($380nm$) to shortwave infrared ($2510nm$). The images are taken over large areas, creating a three-dimensional data cube of radiance, where two dimensions are the spatial domain (i.e., 2D-image) and the third one is in the spectral domain (i.e., wavelength) as shown in Figure 3, which visualizes the relationship between the two datasets. This data is collected in "Four Corner Area" (FCA), the geographical US-Mexico border.

The terrain types include plains, mountain regions, large warehouses, vegetation, water bodies, deeply irrigated fields, livestock farms, coal mines, and other CH_4 emitting areas. The aircraft with AVIRIS-NG instrument flies at a height of $3km$ above the ground. There are multiple CH_4 leakage candidate regions. The ground terrain also contains a large number of confusers in CH_4 detection, for example, paints of hydrocarbons on the roof of warehouses. Paint exhibits similar characteristics to CH_4 and cause strong false positives.

2. Approach

The proposed approach tackles two versions of CH_4 plume dataset. Dataset A is the data pre-processed by Jet Propulsion Laboratory (JPL)[29], where 432-band measurements are processed into one single-channel array using conventional match-filtering techniques with the CH_4 sig-

Figure 4: Sample terrain image (dimensions: 850×1300 pixels) from a flightline (ang20150419t163741). The image has an area of approximately $8 \times 10^5 km^2$. It is reconstructed using the radiance values from the visible spectra ($400nm - 700nm$) from Raw data (χ_B).

nature as the target. The conventional match-filtering technique takes 180 minutes per datapoint to process 432bands into 1 single channel output. The optimized implementation has reduced this processing time to 15 minutes per datapoint. The single channel array is stacked with three other bands, each selected from the visual red, blue, and green wavelengths. The proposed naive single-band solution uses dataset A to evaluate and validate the initial findings and tune a binary plume detector. Dataset B is the original 432-band raw dataset. This dataset is used to design, develop, and evaluate the proposed H-mrcnn solution, which is the formalized naive single-band detector.

Methane Detection as a Segmentation Problem. Gas emitting from a point source has a specific shape texture as it moves through the atmosphere and differs from the underlying terrain. The shape indicates the source or origin of gas, as the gas emitting from a point source has specific plume-like morphology. Mask-RCNN is suited for this problem as it looks for specific patterns in the underlying distribution. In this application it learns the terrain and plume shape, which serves to enhance the detection and eliminate ground terrain confusers(false positives).

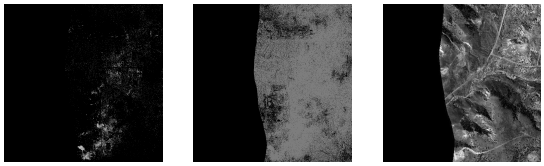
2.1. JPL Dataset A (χ_A)

In dataset A, each data array represents a flightline of the aircraft with the AVIRIS-NG instrument. Visualization of sample image (3-bands, RGB) is shown in Figure 4. The gas plume information is available in the fourth band in the form of ppm×m (part per million per meter) values. The value at each pixel represents the enhancement in CH_4 concentration at that location. The ratio of plume to image pixel counts is very small (i.e., small-pixel footprint).

Data Processing Pipeline. There are only 46 data points ($22000 \times 1400 \times 4$) available to train. The neural network is trained to generate segmentation map of the methane plume and eliminate the false positives. The pre-process converts the single band with CH_4 information into two - the first band is data point level normalization (local normalization) and second band is whole dataset (46 data points) level normalization (global normalization). The data is appended to ground terrain information (greyscale) as the third band. The local normalized band provides precise plume boundaries, see Figure 5(a). The white pixels represent plume and black pixels are background. The global normalized band provides the network with information about the range of methane signal strength across the whole dataset, see Figure 5(b). The greyscale image provides terrain information to the neural network, see Figure 5(c). Each processed data point dimension is 22000 (rows) × 1400 (cols) ×3 (channels). The processed data points are tiled with sliding window in the spatial domain following sizes:

1. $1024 \times 1024 \times 3$ with stride 512
2. $512 \times 512 \times 3$ with stride 256
3. $256 \times 256 \times 3$ with stride 128
4. $128 \times 128 \times 3$ with stride 64

The sample tiles of size 512×512 are shown in Figure 5 for band-1 (a), band-2 (b), and band-3 (c).



(a) (b) (c)

Figure 5: Components of the input 3-channel images to train and test the naive single-band methane plume detector, where (a) Visualization of locally normalized pixel-intensities, (b) Visualization of globally normalized pixel-intensities, and (c) Visualization of greyscale terrain.

Annotation Generation and Data Augmentation. The annotation is available in dataset A. Training is only done on image tiles, which have plume (the original image only has a very small plume, as shown in Figure 2). The fine-tuning process leverages the built-in data augmentation.

Fine-Tuning Mask R-CNN. The Naive Mask-RCNN is a binary and fine-tuned plume detector. Its output is a segmentation mask of CH_4. The Mask-RCNN detector uses ResNet-101 as the backbone. It builds a feature pyramid and then a region proposal network (RPN) proposes regions of objects (plume). Then, these proposals along with the feature pyramid are used by another neural network that produce mask (plume shape), class point), bounding boxes for each instance of objects (plume). For more details on the architecture, please refer to paper [14]. The default confidence value for each predicted plume is 0.7 [1]. One mask is predicted for each class. A sample prediction of plume and its shape are shown in Figure 7 (a) & (c) for the predicted mask (red:black) and (b) & (d) (red:terrain) for the prediction of methane overlayed on the terrain.

2.2. Raw Hyperspectral Dataset B (χ_B)

Matched filtering is a technique to differentiate between the regions of interest and background pixels. In this case,

the signal of interest is CH_4 and the background is the ground terrain. Let $\mathbf{x}_B \in \mathbb{R}^d$ be a sample from dataset χ_B representing the background terrain pixel, where ($\mathbf{x}_B$ is a signal vector in the spectral domain and where each pixel intensity value is the radiance value at a particular wavelength). The spectrum is represented by $\xi(\mathbf{x}_B)$, when the gas is present. The linear matched filter is modeled as additive perturbation:

$$\xi(\mathbf{x}_B) = \mathbf{x}_B + \epsilon\mathbf{t},$$

where $\mathbf{t}$ is the gas spectrum or target signature and ϵ represents the chemical properties of the gas.
The matched filter is a vector $\alpha \in \mathbb{R}^d$ and the output of the matched filter is the scalar $\alpha^T\mathbf{x}_B$. The objective is to find a filter α such that $\alpha^T\xi(\mathbf{x}_B)$ is different from $\alpha^T\mathbf{x}_B$.

The methane gas spectrum, represented by $\mathbf{t}$ and $\alpha^T\mathbf{t}$, is the matched filter output. The terrain pixel vector and matched filter output are represented by $\mathbf{x}_B$ and $\alpha^T\mathbf{x}_B$, respectively. The average Gas-to-Terrain-Ratio (GTR) is:

$$\text{GTR} = \frac{|\alpha^T\mathbf{t}|^2}{Var(\alpha^T\mathbf{x}_B)},$$

where Var is the variance given by:

$$Var(\alpha^T\mathbf{x}_B) = \langle(\alpha^T\mathbf{x}_B - \alpha^T\boldsymbol{\mu})^2\rangle = \alpha^T\mathbf{K}\alpha,$$

with covariance $\mathbf{K}$ and mean μ.

The magnitude of α does not affect GTR; therefore, optimizing GTR means maximizing $\alpha^T\mathbf{t}$ subject to the constant constraint $\alpha^T\mathbf{K}\alpha = 1$. The Lagrangian multiplier λ is added to loss function *l*:

$$l(\alpha; \lambda, \mathbf{t}, \mathbf{K}) = -\alpha^T\mathbf{t} + \lambda(\alpha^T\mathbf{K}\alpha - 1).$$

Minimizing loss function *l* means maximizing GTR [1]. The loss is simplified by assuming $\mathbf{u}$ is:

$$\mathbf{u} = 2\lambda\mathbf{K}^{1/2}\alpha - \mathbf{K}^{-1/2}\mathbf{t}.$$

Then the loss function *l* is re-written as:

$$l(\alpha; \lambda, \mathbf{t}, \mathbf{K}) = \frac{1}{4\lambda}(\mathbf{u^T u} - \mathbf{t^T K^{-1} t}) - \lambda.$$

The GTR is maximized when $\mathbf{u} = 0$ [27], which yields:

$$\alpha = \frac{\mathbf{K}^{-1}\mathbf{t}}{\sqrt{\mathbf{t}^T\mathbf{K}^{-1}\mathbf{t}}}.$$

Intuitively, the methane gas retrieval exploits the linearized matched filter, where the background terrain radiance is modeled as a multivariate Gaussian with mean $\boldsymbol{\mu}$ and covariance $\mathbf{K}$. The matched filter tests the terrain without gas ($\boldsymbol{H_0}$) against the alternative ($\boldsymbol{H_1}$), where the terrain undergoes a linear perturbation by a gas signature $\mathbf{t}$ via:

$$\boldsymbol{H_0} : \mathbf{x}_B \sim \mathcal{N}(\mu, \mathbf{K}),\ \boldsymbol{H_1} : \mathbf{x}_B \sim \mathcal{N}(\mu + \alpha\mathbf{t}, \mathbf{K}).$$

In the CH_4 case, α is a matched filter vector. The column-wise background estimation assumes that most background pixels do not contain the CH_4 gas plume. Therefore, the optimal discriminant matched filter [27] is:

$$\hat{\alpha}(\mathbf{x}_B) = \frac{(\mathbf{x}_B - \mu)^T\mathbf{K}^{-1}\mathbf{t}}{\sqrt{\mathbf{t}^T\mathbf{K}^{-1}\mathbf{t}}}.$$

The target signal $\mathbf{t}$ represents the change in radiance units of the background caused by adding a unit mixing ratio length of CH_4 absorption. This method uses one uniform $\mathbf{t}$ and does not require computing one for each data point. It is applied to dataset B along with the matched filter and neural network detector modules. The main benefits of this approach include the ability to obtain maximum information about the terrain, omit false positives, and achieve accurate plume contours and shapes as demonstrated in § 3.

The sliding-window approach is used to sample the spectral domain. This approach extracts maximum available information about both the plume and the ground terrain across the available wavelength range ($380nm$ to $2510nm$).

Data Processing Pipeline. Added benefits of using a matched filter along with a neural network architecture is the ability to process the data in its raw form. Each data point is of size 22000 (rows) $\times$598 (cols) $\times$432 (bands/wavelength). This means that the files are massive in storage size ranging from 45GB $\sim$ 55GB per file, which becomes a challenge. The raw data is not ortho-rectified, but it is processed using an in-house optimized matched filter over a set of bands. A sample data point $\mathbf{x}_{Bi} \in \mathbf{x}_B$ with single-band matched filter output given by:

$$\hat{\alpha}(\mathbf{x}_{Bi}) = \frac{(\mathbf{x}_{Bi} - \mu)^T\mathbf{K}^{-1}\mathbf{t}}{\sqrt{\mathbf{t}^T\mathbf{K}^{-1}\mathbf{t}}}.$$

The data is processed by sliding a window along the spectral domain with various window sizes and 50% stride. The input data to the matched filter stage is 22000 $\times$ 598$\times$window-size. This yields 22000 $\times$ 598 $\times$ 1 (i.e., $\hat{\alpha}(\mathbf{x}_{Bi})$). This output is processed as described in § 2.1. The 3-band output (1-band: local normalized, 2-band: global normalized and 3-band: greyscale terrain) is tiled using a sliding along the spatial domain, which is the input to Mask-RCNN. The solution is evaluated using the following:

1. window-size of 200 bands
2. window-size of 100 bands
3. window-size of 50 band

[1] see appendix B for more details

2.3. Ensemble Processing Methods

This section describes the algorithms for match filtering, spatial and spectral sliding-window, and fine-tuning DNN detectors using hyperspectral data for plume representation and detection. The multi-band match filtering process of the 432-band hyperspectral data is detailed in Algorithm 1.

Data: χ_B
Result: ortho-corrected matched filter output
($\hat{\alpha}_{ort}(\mathbf{x}_{Bi})$); i=0 initialization;
for x_B *in* X_B **do**
  create memory map x_B;
  while ***i*** *less then Bands* **do**
    read(x_{Bi}) from band **i** to **i+window-size**;
    for *cols in* x_{Bi} **do**
      Compute **K** and μ of x_{Bi};
      $\hat{\alpha}(\mathbf{x}_{Bi}) = \frac{(\mathbf{x}_{Bi}-\mu)^T\mathbf{K}^{-1}\mathbf{t}}{\sqrt{\mathbf{t}^T\mathbf{K}^{-1}\mathbf{t}}}$;
    end
    $\hat{\alpha}_{ort}(\mathbf{x}_{Bi})$ =
     mapping_to_ortho-corrected_values($\hat{\alpha}(\mathbf{x}_{Bi})$);
    i = i-stride;
  end
end

Algorithm 1: Band-wise matched filter.

The output from the matched filter is tiled to deal with the small-pixel footprint nature of the gas plumes in large-area hyperspectral overhead imagery using Algorithm 2 to produce sliding spatial and spectral data tiles.

Data: $\hat{\alpha}_{ort}(\mathbf{x}_{Bi})$, dimension : $22000 \times 1500 \times 1$
Result: $input\_image_{jj}$
initialization;
for *each file in* $\hat{\alpha}_{ort}(\boldsymbol{x}_{Bi})$ **do**
  $\hat{\alpha}_{ort}(\mathbf{x}_{Bi})_{jL}$ = local normalization(jL)
  $\hat{\alpha}_{ort}(\mathbf{x}_{Bi})_{jG}$ = global normalization(jG)
  $gray(\mathbf{x}_{Bi})_{jT}$ = greyscale terrain(jT);
  Stack (jL, jG, jT) together and create tiles of size
   $tilesize \times tilesize \times 3$;
end

Algorithm 2: Data pre-processing.

The spatial and spectral tiles are used to train and fine-tune an ensemble of weak detectors using Algorithm 3.

Ensemble Mask-RCNN. The processed output for each set of bands is used to train a set of neural networks, we call it Ensemble Mask-RCNN. Each neural network is learning about different set of features, for example: bands sets in the short infrared wavelength region ($2200nm$ to $2500nm$) have more information about the presence of CH_4. Recall that the initial spectral band ($400nm$ to $700nm$) sets have more information about terrain and that the matched filter output is pixel-wise independent; therefore, Mask-RCNN learns about the correlation between pixels that contain gas

Data: $input\_image_{jj}$
Result: binary mask of plume shape
initialization;
for *i in sets(0-50, 25-75, 50-100.....)* **do**
  batch_size = 1;
  learning rate = 0.0001;
  epochs = 50;
  image_per_gpu = 1;
  detection_min_confidence = 0.7;
  load_ground_truth(*.png) refer [2.1];
  load_images($input\_image_{jj}$);
  trained_weights = model.train(weights, images,
   ground_truth);
end
binary_mask = model.predict(trained_weights, $image_{jj}$);

Algorithm 3: Mask-RCNN training and fine-tuning.

information. This information is used to fine-tune one detector (Mask-RCNN) for each set of bands. The output from each detector produces a prediction about the plume shape. The weak decisions are fed to a simple 2-layer perceptron network to learn the contribution of each detector in the ensemble and output a final estimate (as a weighted average sum) of the plumes.

3. Experimental Results

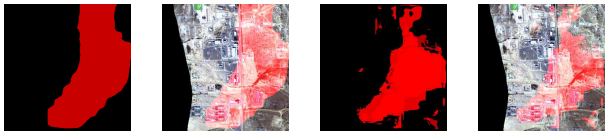
The data is pre-processed on a machine with 16GB RAM and 12 CPU cores. Each image in dataset A is 1∼2 GB in size and processing all of the 46 data points in the dataset takes approximately 45 minutes. The neural network fine-tuning is using one Nvidia GTX 1080Ti GPU. The Mask-RCNN is initialized with MS-COCO [20] weights. The fine-tuning process takes about two hours. A sample measurement of the terrain and the manually generated CH_4 mask (ground truth data) are shown in Figure 6.

(a) (b) (c)

Figure 6: Visualization of a randomly selected datapoint. From left-to-right the images are: the terrain (a), the manually created CH_4 mask (b), and the mask overlayed on the terrain (c).

The ground truth data file is generated by an expert analyst who inspects each CH_4 flightline and manually delineates plumes and separates them from any non-plume artifacts. The manual approach is effective but does not scale, since processing time and human work-hours are significant performance bottlenecks.

The following experiment compares the naive plume detector with the ensemble detector. Qualitative results of the ensemble based plume detection are shown in Figure 7 for the observations collected from terrain shown in (a), with the ensemble predictions shown in (c).



(a) (b) (c) (d)

Figure 7: CH_4 detection output. The naive single-band mask-rcnn detector is shown in (a) as a binary mask (plume vs. no plume) and the detection overlayed on the terrain is shown in (b). The ensemble H-mrcnn detector showing the contour mask of the predicted CH_4 presence is shown in (c). The mask overlayed weight is a concentration analog and the predicted mask overlayed on the overhead terrain image is shown in (d).

Baseline methods. Multiple traditional machine learning approaches that are known to work well with target detection problems [11] and state of the art deep learning image segmentation method [15] are tested on the same dataset as the proposed methods. Logistic regression is commonly used for land cover classification, multinomial logistic regression (MLR) [18] is trained on dataset A, the model performed poorly with IOU of just 5%. Support Vector Machines (SVM) have been successfully applied to hyperspectral data analysis for target detection task. We trained a SVM classifier inspired by [23, 26] on dataset B. For detecting such tiny plumes, SVM performed with 25% higher IOU than logistic regression. The poor performance of SVM is due to the high rate of false positives detection. Gaussian mixture models (GMMs) are also highly infected by false positives. The evaluation is done in using the complete 432 bands data. PCA on raw data followed by logistic regression. This results in poor plume detection, which is caused by the target to image ratio. Capturing 0.80, 0.85, and 0.90 variance did not result in fully getting the target variance into account. Meanshift with watershed algorithm is highly influenced by the terrain and ignores the target.

The performance of H-mrcnn is compared against the inhoused implementation of "Learn to Segment Everything" image segmentation algorithm [15]. This method outperforms the classical machine learning algorithms by overall 19% IOU. Results indicate that the method is capable of eliminating more false positives than the classical methods. However, the method is limited by number of training samples. Unlike in the H-mrcnn solutions, which uses the limited training samples to effectively detect gas plume shape with an average 0.87 IOU.

Evaluation Metrics. The performance metrics for plume detection are designed with the unbalanced and rarity nature of small-pixel footprint plumes in large-area overhead hyperspectral imagery. The metrics include intersection-over-union (IOU), which measures how predicted masks match the manual mask; the euclidean distance between plume centroids which represents how well the plume core is predicted; and the F1-score. The results include data sizes (in number of tiles or detectors) to provide contrast between accuracy and complexity. The overall processing time of (180+7)minutes each datapoint: Dataset B processed by (JPL) conventional matched filter to Dataset A and then Mask-RCNN training, is reduced to (15+7) minutes by H-mrcnn with an IOU increase from 0.832 to 0.879.

PERFORMANCE COMPARISON WITH EXISTING METHODS						
Category	**Method**	**Precision**	**Recall**	**IOU**	**F1**	**Time to Train**
Classical Machine Learning	LogReg [18]	0.52	0.06	0.05	0.11	30hrs CPU
	SVM [23, 26]	0.92	0.3	0.29	.45	39hrs CPU
	GMMs [10]	0.83	0.27	0.4	0.4	20hrs CPU
	Watershed [3]	0.52	0.23	0.18	0.31	21hrs CPU
432 Bands	PCA (.85) + LogReg	0.44	0.07	0.06	0.12	40hrs CPU
	PCA (.85) + SVM	0.84	0.33	0.31	0.47	70hrs CPU
Deep Learning	Segment Everything [15]	0.8	0.55	0.48	0.65	25hrs GPU
	H-mrcnn (proposed)	0.96	0.91	0.87	0.94	30hrs GPU

Table 1: Performance comparison of the proposed H-mrcnn method

Plume Overlap. The performance of the proposed solution is evaluated based on the ratio of predicted plume area overlapping with the human generated ground truth plume area. The plume overlap provides a quantitative measure of how well the proposed naive single-band method performs on predicting the plume shape and location. The higher the plume overlap, the better the prediction. The results summarized in Table 2 show that the naive single-band detector achieves maximum overlap at spatial pixel dimensions of 256×256 with a stride of 128.

Distance between Centroids. The second metric is the Euclidean distance between the centroid of the human generated ground truth plume and the predicted plume. The centroid coordinates are the arithmetic mean position of all the points in the plume shape. The smaller the distance better the prediction. As show in Table 2 the predicted plumes and ground truth coincide best with spatial dimension are 256×256 with a stride of 128.

Plume Detection and Overlap. The following experiments validate the use and design of the DNN-based naive and ensemble detectors. The performance of the naive single-band detector is shown in Table 2. The best performance is achieved at spatial dimensions of 256×256 pixels a stride of 128 pixels. At this resolution, the network is capable of properly representing the mask. The performance

decreases with spatial dimension 128×128 pixels and stride of 64, this is because the whole tile is a plume/mask. At this ratio there is not enough information to effectively represent and learn the background. This results serves to design the individual elements of the ensemble of detectors, where each detector is tuned for specific spectral region using spatial 256×256 pixels tiles and 128 stride.

Naive Mask-RCNN (Dataset A)					
Spatial Dimension	**IOU**	**F1-score**	**Distance**	**Time**	**#Tiles**
1024x1024 (overlap 512)	0.590	0.742	23.96	90	150 (4 : 6)
512x512 (overlap 256)	0.769	0.869	13.24	150	500 (15 : 20)
256x256 (overlap 128)	**0.832**	**0.923**	**5.83**	**270**	200 (35 : 40)
128x128 (overlap 64)	0.762	0.865	22.36	420	7000 (100 : 125)

Table 2: Naive Mask-RCNN detection. Performance metrics for the tile spatial dimensions for various pixel (window size) combinations with 50% overlap (stride). The metrics include IOU: intersection-over-union for mask and ground-truth overlap; F1-score: plume detection performance for unbalanced plume vs. no-plume datapoints; Distance: the estimated centroid euclidean distance (plume's first geometric moment); Time: is the approximate processing time in minutes; the number of tiles generated by each configuration is shown in # Tiles and (background:plume) ratio. Time excludes processing time by JPL from Dataset B to Dataset A

Ensemble of Detectors. A bank of detectors predicts plume shape. As shown in Table 3. The model best performs when the set size in spectral domain is 50 bands in a set with stride 25. The spatial dimensions are 256×256 with stride 128, which were learned from the naive detector. Using a spectral set of 50 allows the network to cover the visible spectrum, i.e., $380nm$ to $650nm$ and learn about the terrain. This information is used to reduce false positives. In addition, the ensemble uses individual wavelength neighborhoods. The balance is critical as the number of detectors and processing time increases exponentially making the system inefficient. The final plume estimated is produced by 2-layer perceptron, which decides the weights for each detector in detector bank. As shown in Table 3 row 1, on assigning equal weights to all the detectors, the final predicted plumes have smaller IOU values. Using a network to learn and estimate decision weights for each member in the ensemble produces an accurate plume prediction.

H-MRCNN (Dataset B)					
Spectral Dimension	**IOU**	**F1-score**	**Distance**	**Time**	**Bank**
200 bands (overlap 100)	0.645	0.772	16.71	$90^*/90^+$	4
100 bands (overlap 50)	0.706	0.814	4.24	$810^*/90^+$	9
50 bands (overlap 25)	**0.854**	**0.921**	**4.12**	$1500^*/90^+$	**17**

Table 3: Bank of Detectors showing performance metrics for the tile spectral dimensions for various band (window size) combinations with 50% overlap (stride). The metrics include IOU: intersection-over-union for mask and ground-truth overlap; F1-score: plume detection performance for unbalanced plume vs. no-plume datapoints; Distance: the estimated centroid euclidean distance (plume's first geometric moment); Time: is the approximate processing time in minutes(includes processing of Raw data), where the symbols * and + represent not-prarallelized and parallelized processes, respectively; Bank is the number of detectors generated by each configuration.

H-MRCNN with & without Ensemble Network			
Ensemble Detection	**IOU**	**F1-score**	**Distance**
Uniform Weight for all Predictions	0.854	0.920	4.120
2-Layer Perceptron (range: -1 to 1)	**0.880**	**0.945**	**4.120**

Table 4: Ensemble Detection Performance based on decision fusion mechanism (Uniform vs. 2-layer perceptron). The metrics reported include Intersection-over-Union between the true-labeled and predicted masks (IOU); F1-score for binary incidence and detection of plumes; and Distance: Euclidean distance between the centroids of true plume mask and the predicted plume mask in pixels.)

4. Summary

This work introduces techniques that leverage pre-existing deep neural network based detectors to formulate a naive single-band detector. Also, this work further develops the findings from the design and evaluation of the naive detector (i.e., data sampling parameters). It integrates spectral sampling along with a newly improved and optimized match-filtering algorithm to process large-area hyperspectral data. The processed hyperspectral data is used to fine-tune and construct an ensemble of detectors. Thorough experimental results indicate that the naive detector matches the performance of human annotations by 83.2% for binary detections. The ensemble approach outperforms the detection of the naive detector by 87.95% and better represents the plume using the ensemble detections. The presented solutions will enable the rapid processing and analysis of gas plumes, removes the confusers (false positives), which produces actionable information and response plans about greenhouse gases in the atmosphere.

Future Work. Future work includes extensions of the H-mrcnn to other gases. In addition, potential future directions include online learning and tuning for methane point sources and diffused sources. As more data becomes available, the study will focus on time-series analysis of plumes (e.g., plume dispersion, gas flux, life-span, and evolution).

5. Acknowledgements

This project was founded in part by The Aerospace Corporation under PO #4600006535. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of The Aerospace Corporation. The authors thank Ronald Scrofano from The Aerospace Corporation for his valuable feedback during this project.

References

[1] W. Abdulla. Mask r-cnn for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask_RCNN, 2017.

[2] A. K. Ayasse, A. K. Thorpe, D. A. Roberts, C. C. Funk, P. E. Dennison, C. Frankenberg, A. Steffke, and A. 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.

[3] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[4] G. Camps-Valls. Support vector machines in remote sensing: the tricks of the trade. In *Image and Signal Processing for Remote Sensing XVII*, volume 8180, page 81800B. International Society for Optics and Photonics, 2011.

[5] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):6232–6251, 2016.

[6] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected topics in applied earth observations and remote sensing*, 7(6):2094–2107, 2014.

[7] Y. Chen, X. Zhao, and X. Jia. Spectral–spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2381–2392, 2015.

[8] Q. Cheng, P. K. Varshney, and M. 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.

[9] 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*, 5(1):9–22, 2005.

[10] C. Frankenberg, A. K. Thorpe, D. R. Thompson, G. Hulley, E. A. Kort, N. Vance, J. Borchardt, T. Krings, K. Gerilowski, C. 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.

[11] U. B. Gewali, S. T. Monteiro, and E. Saber. Machine learning based hyperspectral image analysis: a survey. *arXiv preprint arXiv:1802.08701*, 2018.

[12] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris). *Remote sensing of environment*, 65(3):227–248, 1998.

[13] L. Hamlin, R. 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. IEEE, 2011.

[14] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[15] R. Hu, P. Dollar, K. He, T. Darrell, and R. Girshick. Learning to segment every thing. pages 4233–4241, 2018.

[16] G.-B. Huang, D. H. Wang, and Y. Lan. Extreme learning machines: a survey. *International journal of machine learning and cybernetics*, 2(2):107–122, 2011.

[17] B. ISAC. Differential optical absorption spectroscopy (doas). 1994.

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

[19] S. Kirschke, P. Bousquet, P. Ciais, M. Saunois, J. G. Canadell, E. J. Dlugokencky, P. Bergamaschi, D. Bergmann, D. R. Blake, L. Bruhwiler, et al. Three decades of global methane sources and sinks. *Nature geoscience*, 6(10):813, 2013.

[20] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[21] S. T. Monteiro, Y. Minekawa, Y. Kosugi, T. Akazawa, and K. Oda. Prediction of sweetness and amino acid content in soybean crops from hyperspectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(1):2–12, 2007.

[22] D. A. Roberts, E. S. Bradley, R. Cheung, I. Leifer, P. E. Dennison, and J. S. Margolis. Mapping methane emissions from a marine geological seep source using imaging spectrometry. *Remote Sensing of Environment*, 114(3):592–606, 2010.

[23] W. Sakla, A. Chan, J. Ji, and A. Sakla. An svdd-based algorithm for target detection in hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 8(2):384–388, 2010.

[24] C. Shah, P. K. Varshney, and M. Arora. Ica mixture model algorithm for unsupervised classification of remote sensing imagery. *International Journal of Remote Sensing*, 28(8):1711–1731, 2007.

[25] T. Stocker. *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2014.

[26] D. M. Tax and R. P. Duin. Support vector data description. *Machine learning*, 54(1):45–66, 2004.

[27] J. Theiler, B. R. Foy, and A. 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, page 656503. International Society for Optics and Photonics, 2007.

[28] D. Thompson, I. Leifer, H. Bovensmann, M. Eastwood, M. Fladeland, C. Frankenberg, K. Gerilowski, R. 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.

[29] D. R. Thompson, A. Karpatne, I. Ebert-Uphoff, C. Frankenberg, A. K. Thorpe, B. D. Bue, and R. O. Green. Isgeo dataset jpl-ch4-detection-2017-v1. 0: A benchmark for methane source detection from imaging spectrometer data. 2017.

[30] A. K. Thorpe, C. Frankenberg, D. R. Thompson, R. M. Duren, A. D. Aubrey, B. D. Bue, R. O. Green, K. Gerilowski, T. Krings, J. 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, 2017.

[31] A. K. Thorpe, D. A. Roberts, E. S. Bradley, C. C. Funk, P. E. Dennison, and I. 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.

[32] J. Xia, P. Du, X. He, and J. Chanussot. Hyperspectral remote sensing image classification based on rotation forest. *IEEE Geoscience and Remote Sensing Letters*, 11(1):239–243, 2013.

[33] W. Zhao and S. Du. Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8):4544–4554, 2016.

[34] W. Zhao, Z. Guo, J. Yue, X. Zhang, and L. Luo. On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery. *International Journal of Remote Sensing*, 36(13):3368–3379, 2015.