Supplementary Materials : 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

APPENDIX A: Additional Data Information

0.1. Dataset A.

The dataset consists of flight-lines from the "Four Corners Campaign" survey. It is split into flight-lines representing individual contiguous images strips acquired by the AVIRIS-NG instrument as it flew over the terrain. AVIRIS-NG is a push-broom imaging spectrometer with 598 across track elements of $1\ m$rad instantaneous field of view (FOV) covering a total FOV of 34° . AVIRIS-NG measures reflected solar radiance at the nadir viewing geometry at a $100Hz$ readout rate across 432 channels between $380nm$ and $2510nm$ with a spectral sampling of $5nm$. Each flight-line is associated with three files-two files provides the meta data and one file represents the manually identified and labeled location of the Ch_4 plume.

1. **A header file that provides META data, with a filename suffix * cmf_v1f_img.hdr :** This file describes the specific dimensions and layout of the flight-line, and embeds geographic projection information. This files allows to read the data from second input file
2. **The imaging spectrometer data, with a filename suffix * cmf_v1f_img :** A binary file representing the CH_4 image data, with four channels:
   (a) Radiance in visible red wavelengths
   (b) Radiance in visible green wavelengths
   (c) Radiance in visible blue wavelengths
   (d) Estimated CH_4 enhancement above the background in ppm(parts per million meters). This keeps the image in rectangular shape with the pheriphery filled with "-9999.0" values.

   The data is in Band Interleaved by Line (BIL) ordering. BIL ordering signifies the 3D matrix is indexed first by image row, channel and image column. The data is 32 bit IEEE little endian floating point format.
3. **The label file from manual source analysis, with a filename suffixed *.png :** An image file in .png format, representing a human labeling of the methane source in the flightline. It has three channels with values in $[0 - 255]$ range, code according to : Red (255,0,0) : plume, associated with a point source. Blue (0,0,255): plume, associated with a diffused source. Black (0,0,0): no plume (or unlabeled).

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

0.2. JPL Creation of Dataset A (χ_A)

In dataset A, each data array represents a flight-line of the aircraft with the AVIRIS-NG instrument. The *cmf_v1f_img image file holds 4-bands data in ENVI format. Dimensions of the *cmf_v1f_img are in range 22000 pixels $\times$ 1400 pixels $\times$ 4 bands. The gas plume information is available in the fourth band in the form of ppm$\times$km (part per million per meter) values. The value at each pixel represents the amount CH_4 at that location. The ratio of plume to image pixels counts is very small (i.e., small-pixel footprint challenge), which addressed by the proposed data pre-processing and sampling method.

0.3. Dataset B.

The data consists of flightlines from the Four Corners campaign. It contains approximately 700 flightlines, where only 46 human-annotated data are available for this study. Each flight-line is associated with the following four files:

1. **A header file the provides the META data, with the filename suffix * rdn_v1f_clip.hdr :** describes the specific dimensions and layout of the flight-line, and embeds geographic projection information. It enables reading data from the corresponding image file.
2. **Imaging spectrometer data (suffix * rdn_v1f_clip):** A binary file, representing the radiance values(units $\mu W cm^{-2} nm^{-1} sr^{-1}$) at each wavelengths. The file has 432 channels, each channel represent the wavelength the image is captured by the imaging spectrometer with spectral resolution of $5nm$, starting from channel 0 at wavelength $380nm$ goes up to channel

431 at wavelength $2510nm$. Each of the image files are of approximately 50 Gigabytes.

3. **Metadata Header (suffix * rdn_glt.hdr):** This file describes dimensions and serves as header file to read the corresponding glt image file.
4. **GPS data (name suffix * rdn_glt):** This file has the map of geo-reference co-ordinates of each and every pixel of * rdn_v1f_clip file. This file is used to ortho-correct the processed match-filtered output and reduce the image distortion caused by aircraft.

APPENDIX B: Matched Filter

0.4. Matched Filter Derivation

The output of the H-mrcnn implementation of the matched filter is a representation of CH_4 signal strength. It assumes that the observed spectrum of background pixel is $\mathbf{x}_B \in \mathbb{R}^d$, then $\xi(\mathbf{x}_B)$ is the spectrum of that pixel if the signal of interest for CH_4 is added to background pixels. In linear Adaptive matched Filter, the effect of signal is modeled as additive perturbation given by:

$$\xi(\mathbf{x}_B) = \mathbf{x}_B + \epsilon\mathbf{t}, \tag{1}$$

where $\mathbf{t}$ is the signal/gas spectrum or target signature and ϵ represents the chemical properties of the gas.

The matched filter is a vector $\alpha \in \mathbb{R}$ and the output of the matched filter is a scalar value $\alpha^T\mathbf{x}_B$. The filter α becomes $\alpha^T\xi(\mathbf{x}_B)$ and differs from $\alpha^T\mathbf{x}_B$.

If $\mathbf{t}$ is the signal of interest, then $\alpha^T\mathbf{t}$ is the effect of matched filter on the signal. If $\mathbf{x}_B \in \mathbb{R}^d$ is a hyperspectral pixel value without signal, then $\alpha^T\mathbf{x}$ is the effect of the matched filter on the background pixel spectrum. The average Gas-to-Terrain-Ratio (GTR) is given by:

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

where Var is the variance, is computed from:

$$Var(\alpha^T\mathbf{x}_B) = \langle(\alpha^T\mathbf{x}_B - \alpha^T\boldsymbol{\mu})\rangle = \alpha^T\mathbf{K}\alpha, \tag{3}$$

with $\mu = \langle x\rangle$ mean and $\mathbf{K} = \langle(x-\mu)^T(x-\mu)\rangle$ covariance. Since the magnitude of α does not affect the GTR in Eq. 3, so constraint can be imposed on the magnitude of α. This optimizes the GTR in Eq. 2, which is equivalent to maximizing $\alpha^T\mathbf{t}$ subject to the constraint $\alpha^T\mathbf{K}\alpha$. Using $\alpha^T\mathbf{K}\alpha$ = 1, leads to the following Lagrangian formulation:

$$l = -\alpha^T\mathbf{t} + \lambda(\alpha^T\mathbf{K}\alpha - 1) \tag{4}$$

Minimizing the loss function l is equivalent to maximizing the GTR, Let's:

$$\mathbf{u} = 2\lambda\mathbf{K}^{1/2}\alpha - \mathbf{K}^{-1/2}\mathbf{t}, \tag{5}$$

where the scalar value can be written as:

$$\mathbf{u^T u} = (2\lambda\mathbf{K}^{1/2}\alpha - \mathbf{K}^{-1/2}\mathbf{t})^T(2\lambda\mathbf{K}^{1/2}\alpha - \mathbf{K}^{-1/2}\mathbf{t}) \tag{6}$$

$$= 4\lambda^2\alpha^T\mathbf{K}\mathbf{t} - 4\lambda\alpha^T\mathbf{t} + \mathbf{t}^T\mathbf{K}^{-1}\mathbf{t} \tag{7}$$

so Eq. 4 becomes:

$$l(\alpha; \lambda, \mathbf{K}, \mathbf{t}) = \frac{1}{4\lambda}\mathbf{u^T u} - \lambda - \frac{1}{4\lambda}\mathbf{t}^T\mathbf{K}^{-1}\mathbf{t} \tag{8}$$

and it is clear that this is minimized when $\mathbf{u} = 0$, from which Eq. 5 implies $\alpha = \frac{1}{2\lambda}\mathbf{K}^{-1}\mathbf{t}$. Setting $\alpha^T\alpha = 1$ yields $\lambda = \frac{1}{2}(\mathbf{t}^T\mathbf{K}^{-1}\mathbf{t})^{-\frac{1}{2}}$ with (final) matched filter response as:

$$\alpha = \frac{\mathbf{K}^{-1}\mathbf{t}}{\sqrt{\mathbf{t}^T\mathbf{K}^{-1}\mathbf{t}}}. \tag{9}$$

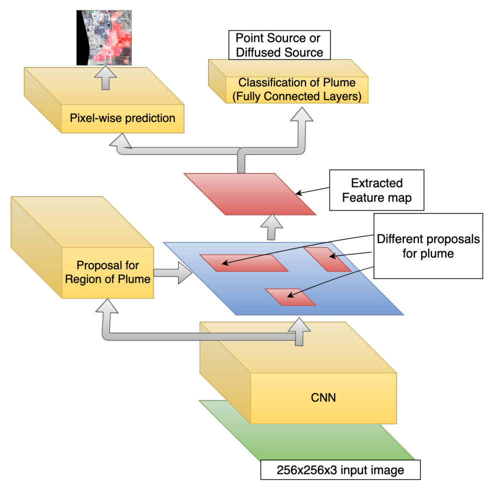

Figure 1: Architecture of Naive Mask-RCNN

0.5. Overview of Mask-RCNN

Mask RCNN has been the new state of art in terms of instance segmentation. There are two stages of Mask RCNN. First, it generates proposals about the regions where there might be an object based on the input image. Second, it predicts the class of the object, unlike regular Mask-RCNN we are not using the bounding box representation of it. Figure 1 shows a simple block diagram architecture of Mask-RCNN model used in Naive detector and H-mrcnn. The Backbone used is Resnet-101 in Naive detector. The final outputs as shown in the Figure 1 are segmentation mask of plume and class.

APPENDIX C: Additional Experimental Results Visualizations

0.6. Distance between Centroids

This section shows sample results for the distance between the centroid locations of the ground truth and the predicted plume as shown in Figure 2, where two black dots inside the plume masks are the computed centroids.

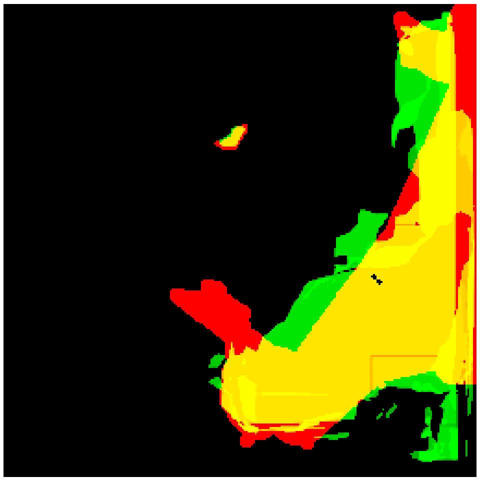

Figure 2: Sample results of distance between the centroids of ground truth and prediction by our model. Red represents the human expert generated ground truth and Green+yellow is the prediction from the network. The 2 black dots represents the centroids of ground truth and prediction and line connecting them is distance between the centroids.

0.7. Predictions by Ensemble method

Sample predictions are shown in Figures 5] and 6, which are the outputs of the Ensemble-Mask-RCNN (Regional Convolution Neural Network). Column 1 is the ground truth generated by human experts. Column 2 is a visualization of ground terrain of the same area where the plume is shown in ground truth. Column 3 is overlay of ground truth on terrain from column 2. Columns 4, 5, and 6 show H-mrcnn predictions with different band-window sizes, each configuration uses a 50% overlap (i.e., stride). Column 4 represents the highest IOU value for plume overlap. Similarly, columns 5 and 6 show results for band-window size of 100 and 200 respectively with stride of 50%. Figure 6 shows a unique example where the ratio of number_of_plume_pixel to number_of_image_pixel is lowest. This means that the plume covers approximately 50 pixels from an image with dimensions 3.08×10^7 pixels, which is used to represent true negative predictions.

Total number of Plumes	Correctly detected Plumes	Not detected plume	Fasle positive
∼80	∼70	∼5	3

Table 1: General Statistics of Detection

0.8. Prediction observation

The total number of plumes in the whole dataset are approximately around 80 in number. By 80 plumes means that one data point can have multiple plumes. The H-mrcnn detected approximately 70 plumes with a good accuracy. The few of the cases where the method fails to detect plume is when the plume size is 10^{-6} times smaller then number of pixel in the image containing the plume. There are only 3 cases of false positives detection. Those are when there is no plume in the whole image and the method detects a plume in that location. Other cases of false positives are as follows. When the plume is very big, then the edges of the big plume have tiny plumes on the sides. These are the cases when the ensemble network computes a weighted sum to predict the final plume. In some cases, when there is plume but because of a wet field on the ground or any water body on the ground, in those cases the magnitude of radiance values is very low in those pixel, in such cases even if the plume is present but the model is not able to completely detect it. This is a drawback of the AVIRIS-NG sensor as it is not able to record the correct radiance values in those areas.

Enhanced Visualization of the Methods. Enhanced result visualizations are shown in Figure 3 and 4 for the naive single-band detector (method 1) and ensemble H-mrcnn (method 2), respectively.

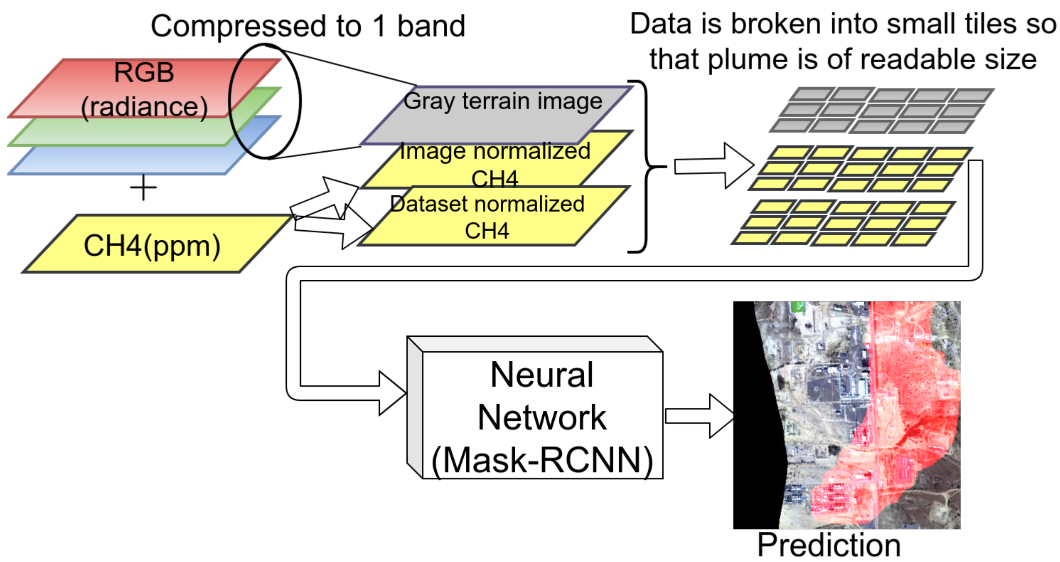

Figure 3: Data munging and fine-tuning processing diagram for the naive single-band and binary plume detector. From top-left to bottom-right: the input arrays (top-left) are locally and globally normalized (top-center). The processed 3-channel array is tiled (top-right) to deal with small-pixel footprint plumes. The tiles are used to fine-tune a naive binary plume detector (bottom-center), which produces a binary estimate of the gas plume overlayed on the observed terrain (bottom-right).

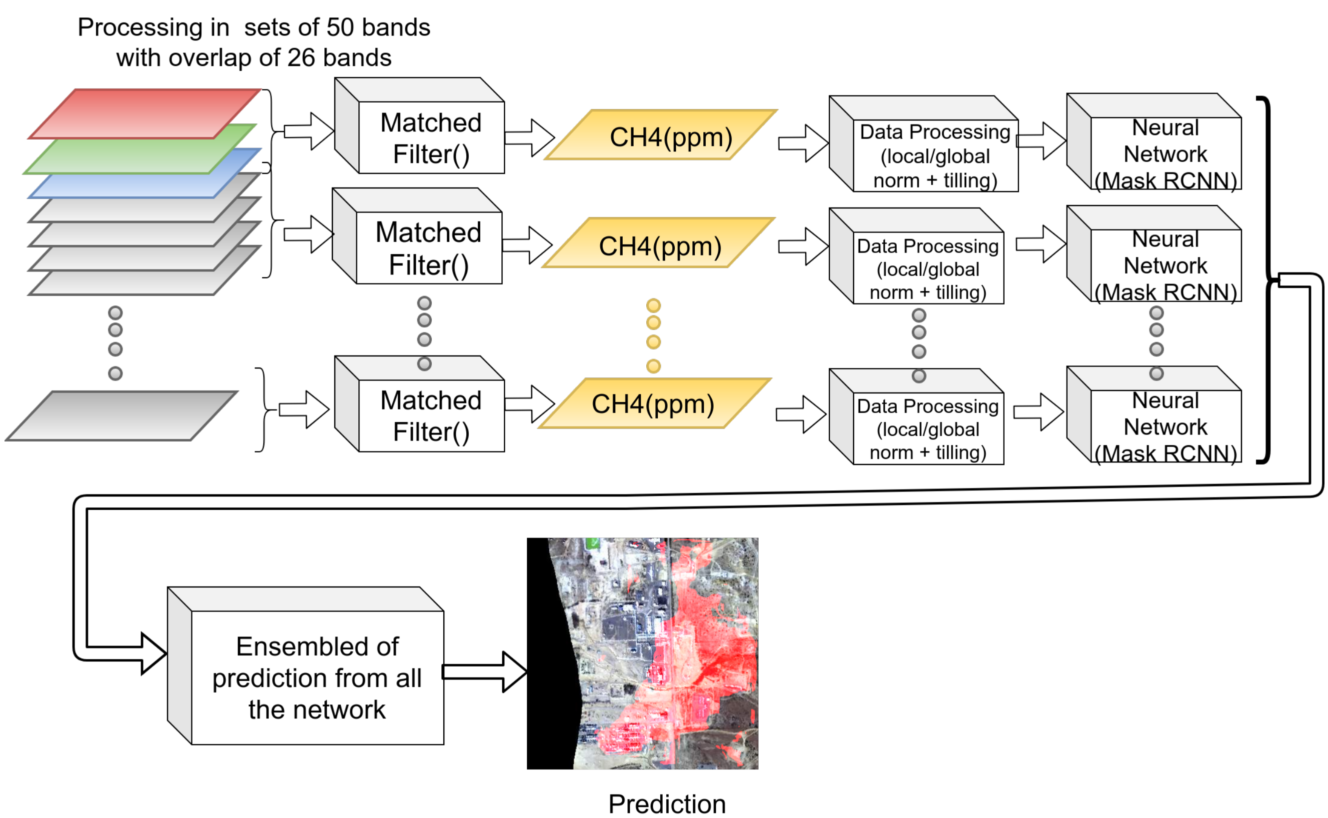

Figure 4: H-mrcnn ensemble structure. Bank of detectors fine-tuned on matched filter outputs. The ensemble network produces better estimates of the plume shape and a concentration analog.

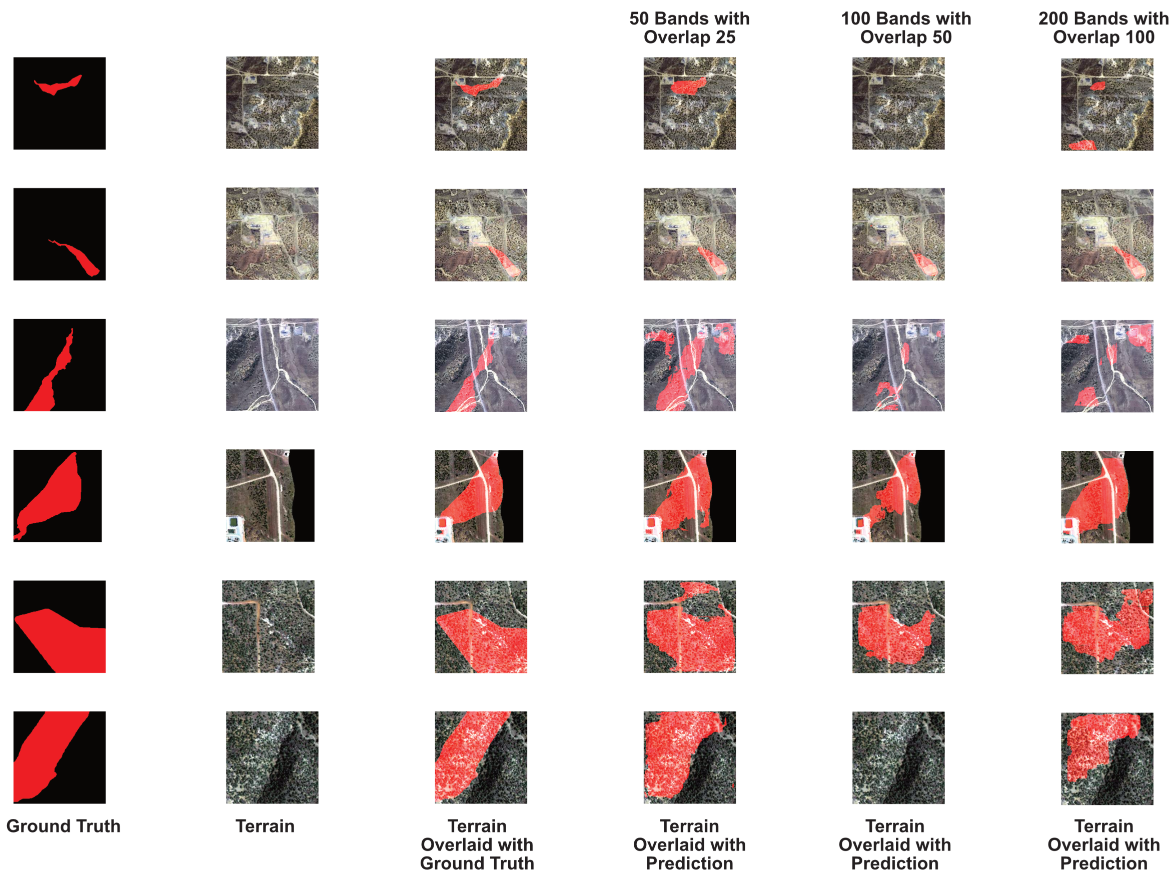

Figure 5: Sample results for Ensemble H-mrcnn. Columns 1, 2, and 3 are ground truth and ground terrain, columns 4, 5, and 6 show the prediction output from the H-mrcnn method. The best results are shown in column 4 for a 50 bands window and 25 bands stride. Note that the predictions are overlaid on the terrain and that each tile has 256×256 pixel dimensions.

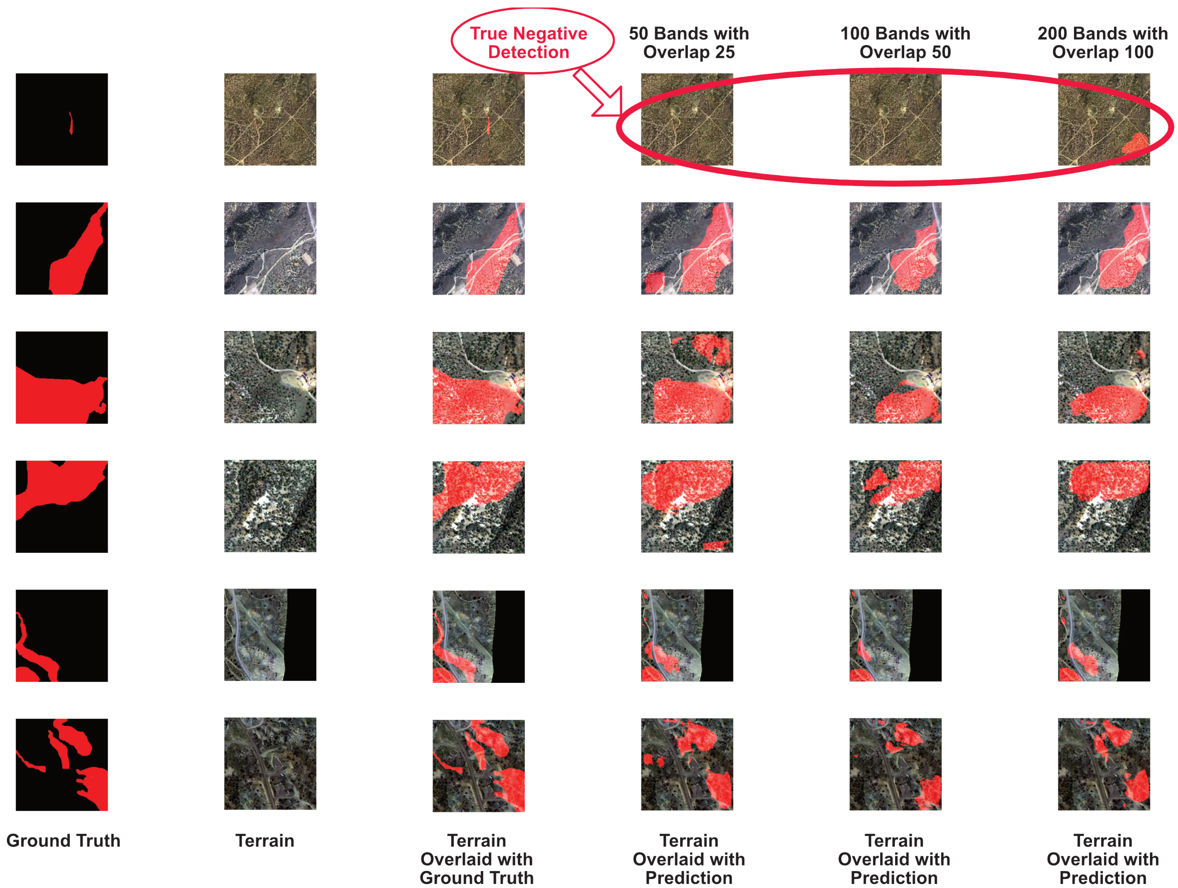

Figure 6: Sample results for Ensemble H-mrcnn. Columns 1, 2, and 3 are ground truth, terrain, and ground on terrain; columns 4, 5, and 6 show the predicted outputs from the H-mrcnn ensemble method. The best results are shown in column 4 for a band-widow of size 50 and stride of 25. Note that the predictions are overlaid on the terrain. Row 1 in columns 4, 5, and 6 (circled in red) depict a true negative detection sample case.