Deep Learning for Clouds and Cloud Shadow Segmentation in Methane Satellite and Airborne Imaging Spectroscopy

Manuel Pérez-Carrasco
Center for Data and AI
Universidad de Concepción
Concepción, Chile
maperezc@inf.udec.cl

Maya Nasr
Environmental Defense Fund
Harvard University
Cambridge, MA 02138, U.S.A.
mayanasr@g.harvard.edu

Sébastien Roche
Environmental Defense Fund
Harvard University
Cambridge, MA 02138, U.S.A.
sroche@g.harvard.edu

Chris Chan Miller
Environmental Defense Fund
Harvard University
Cambridge, MA 02138, U.S.A.
cmiller@g.harvard.edu

Zhan Zhang
John A. Paulson School of
Engineering and Applied Sciences
Harvard University
Cambridge, MA 02138, U.S.A.
zhanzhang@g.harvard.edu

Core Francisco Park
Department of Physics
Harvard University
Cambridge, MA 02138, U.S.A.
corefranciscopark@g.harvard.edu

Eleanor Walker
John A. Paulson School of
Engineering and Applied Sciences
Harvard University
Cambridge, MA 02138, U.S.A.
ewalker@g.harvard.edu

Cecilia Garraffo
AstroAI
Center for Astrophysics | Harvard & Smithsonian
Cambridge, MA 02138, U.S.A.
cgarraffo@cfa.harvard.edu

Douglas Finkbeiner
Department of Physics
Harvard University
Cambridge, MA 02138, U.S.A.
dfinkbeiner@cfa.harvard.edu

Ritesh Gautam
Environmental Defense Fund
New York, NY 10010, U.S.A.
rgautam@edf.org

Steven Wofsy
Department of Earth and Planetary Sciences
Harvard University
Cambridge, MA 02138, U.S.A.
wofsy@g.harvard.edu

Abstract

Effective cloud and cloud shadow detection is a critical prerequisite for accurate retrieval of concentrations of atmospheric methane (CH_4) or other trace gases in hyperspectral remote sensing. This challenge is especially pertinent for Methane-

SAT, a satellite mission launched in March 2024, to fill a significant data gap in terms of resolution, precision and swath between coarse-resolution global mappers and fine-scale point-source imagers of methane, and for its airborne companion mission, MethaneAIR. MethaneSAT delivers hyperspectral data at an intermediate spatial resolution ($\sim 100 \times 400$ m), whereas MethaneAIR provides even finer resolution (~ 25 m), enabling the development of highly detailed maps of concentrations that enable quantification of both the sources and rates of emissions. In this study, we use machine learning methods to address the cloud and cloud shadow detection problem for sensors with these high spatial resolutions. Cloud and cloud shadows in remote sensing data need to be effectively screened out as they bias methane retrievals in remote sensing imagery and impact the quantification of emissions. We deploy and evaluate conventional techniques—including Iterative Logistic Regression (ILR) and Multilayer Perceptron (MLP)—with advanced deep learning architectures, namely U-Net and a Spectral Channel Attention Network (SCAN) method. Our results show that while conventional methods are lightweight, they struggle with spatial coherence and boundary definition, in turn affecting the detection of clouds and cloud shadows. Deep learning models substantially improve detection quality: U-Net performs best in preserving spatial structure, while SCAN excels at capturing fine boundary details. Notably, SCAN surpasses U-Net on MethaneSAT data, underscoring the benefits of incorporating spectral attention for satellite-specific features. Additionally, we combine the predictions of both U-Net and SCAN through a Convolutional Neural Network (CNN). This method achieves state-of-the-art performance on both MethaneAIR (78.50±3.08% F1) and MethaneSAT (78.80±1.28% F1) datasets with efficient inference (4.1 ms per 1,000 km^2). This in-depth assessment of various disparate machine learning techniques, as applied to MethaneSAT and MethaneAIR imaging spectroscopic data at varying spatial resolutions, demonstrates the strengths and effectiveness of advanced deep learning architectures in providing robust, scalable solutions for clouds and cloud shadow screening towards enhancing methane emission quantification capacity of existing and next-generation hyperspectral missions.

Our code is available at: `https://dataverse.harvard.edu/dataverse/MAIR_MSAT_CLOUD_SHADOWS`

1 Introduction

Remote sensing has rapidly evolved as a key tool for quantifying emissions of critically-important greenhouse gases, including carbon dioxide (CO_2) and methane (CH_4). Among these, methane is especially critical: although it has a relatively short atmospheric lifetime of about twelve years, CH_4 exhibits over 80 times the warming potential of CO_2 during the first two decades after emission [1, 2]. This makes methane an attractive target for near-term climate mitigation efforts [3].

Methane mitigation efforts have advanced in the past few years with the Global Methane Pledge [4] signed by over 150 countries aiming to reduce anthropogenic methane emissions by 30% by year 2030. Sectorally, the Oil and Gas Decarbonization Charter (OGDC) has pledged to reduce methane emissions intensity down to 0.2% of production by 2030, which includes over 50 oil and gas companies accounting for more than 40% of global production[5]. It is important to track the performance of critical methane mitigation targets, and remote sensing has emerged as an effective tool for globally monitoring methane emissions at scale.

Two primary remote sensing strategies have emerged for monitoring methane: high-resolution imaging spectrometers, such as AVIRIS-NG, GHGSat, Carbon Mapper, and Sentinel-2, which excel at identifying point sources [6, 7, 8], and low-resolution global mappers like TROPOMI and GOSAT, which offer daily or near-real-time regional coverage [9, 10]. These hyperspectral missions have provided important data on global scale atmospheric methane concentrations and information on high-emitting point sources in targeted domains. However, there exists an observing and data gap at the scales of individual areas where emission quantification information is not readily available for assessing emissions at high spatial resolution, high precision and wider coverage - limiting our overall understanding of methane emission sources and magnitude.

The MethaneSAT mission [11], launched in March 2024, directly addresses this gap by combining fine-scale resolution, high-precision measurements with broad swath coverage of 220 km at nadir (extendable to over 400 km swath at greater off-nadir viewing geometries). The satellite features two imaging spectrometers: one targeting CH_4 (1589–1686 nm) and another measuring O_2 (1249–1305 nm), enabling precise retrieval of methane concentrations over oil and gas production basins and agricultural regions distributed globally. MethaneAIR, its airborne companion [12], provides critical data for algorithm development and instrument validation, as well as ongoing mapping of methane emissions. Figure 1 illustrates the contrasting spatial coverage capabilities of MethaneSAT and MethaneAIR.

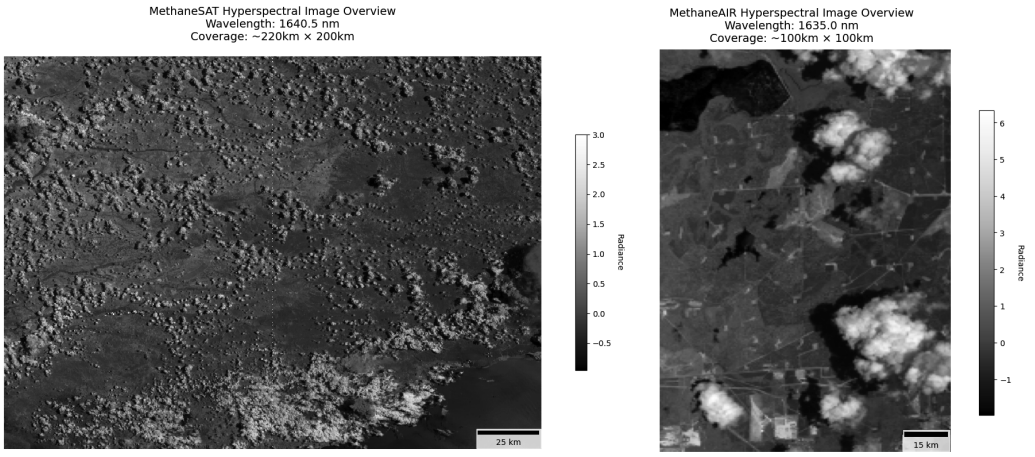


Figure 1: Hyperspectral methane observations from MethaneSAT and MethaneAIR platforms showing typical mapping coverage and spatial extent. **Left**: MethaneSAT image at 1640.5 nm wavelength, captured on September 4 of 2024, covering approximately 220 km × 200 km area with 25 km scale bar. **Right**: MethaneAIR image at 1635.0 nm wavelength, captured on June 2 of 2023, covering approximately 100 km × 100 km area with 15 km scale bar.

A key obstacle in retrieving accurate methane concentrations from hyperspectral imagery is the presence of clouds and cloud shadows, which introduce significant artifacts. When a pixel includes a cloud, surface reflectance is occluded, leading to signal loss that is required for retrieving methane concentrations in cloud-free column atmosphere. In contrast, cloud shadows obscure the solar illumination path, making the optical path length ambiguous. These artifacts impact methane retrievals differently across wavelengths due to variable scattering and absorption in the shortwave infrared (SWIR), resulting in class-specific spectral distortions. As a result, both the spectral and spatial domains encode rich, complementary information about the presence of clouds and shadows—making this problem well-suited to machine learning approaches that can leverage high-dimensional contextual features.

In this work, we investigate a range of machine learning algorithms for the detection of clouds and shadows in MethaneSAT and MethaneAIR hyperspectral imagery. We evaluate conventional models such as logistic regression [13, 14, 15] and multilayer perceptrons (MLPs) [16, 17, 18, 19, 20], as well as deep learning approaches including U-Net [21, 22, 23, 24, 25, 26, 27, 28] and Spectral Channel Attention Network (SCAN), the latter designed to apply channel-wise attention mechanisms for hyperspectral band selection. Our results demonstrate that while conventional models offer computational efficiency, they struggle to resolve spatially complex or spectrally ambiguous regions. Deep networks, on the other hand, yield higher accuracy—U-Net excels in spatial consistency, while SCAN improves spectral boundary delineation, particularly for shadow detection.

To further improve performance, we develop ensemble architectures that fuse predictions from U-Net and SCAN. We explore both MLP- and CNN-based fusion strategies, with the latter achieving up to 8% improvements in F1 score over baseline models. These ensembles effectively combine spatial coherence and spectral attention to outperform individual models, especially in regions with complex topography or partial cloud cover.

Our findings reinforce the importance of robust cloud and shadow masking for hyperspectral CH_4 retrieval and provide scalable solutions to enhance the fidelity of MethaneSAT and MethaneAIR products, that are aimed at advancing the quantification of emissions from individual oil/gas and agricultural regions around the world. More broadly, they offer potential applications and insights to other hyperspectral missions facing similar atmospheric challenges in remote sensing-based greenhouse gas monitoring.

2 Background

Hyperspectral semantic segmentation (HSS), also known as pixel-wise classification, aims to assign distinct labels to each pixel based on their spatial and spectral characteristics [29, 30, 31]. The rich spectral information contained in hyperspectral data, spanning multiple contiguous narrow bands, enables precise discrimination of materials and atmospheric features that may be indistinguishable in conventional imaging.

Conventional approaches for HSS primarily relied on spectral-based classification, where each pixel is classified independently based on its spectral signature [32, 16, 17, 14]. Methods like logistic regression [33, 14, 15] and multilayer perceptrons (MLPs) [16, 17, 18, 19, 20] operate directly on the spectral vectors, learning the complex relationships between spectral bands and target classes. While these spectral-only approaches can effectively leverage the rich spectral information, they fail to capture important spatial context and patterns that could significantly improve segmentation accuracy.

To address this limitation, modern deep learning architectures have been developed to jointly exploit both spectral and spatial information [32, 34, 35]. The U-Net [22, 21] architecture, originally designed for biomedical image segmentation [22], has been adapted for multispectral cloud and shadow segmentation using Landsat 8 [23], Sentinel-2 [24, 25, 26], MODIS [27], and Gaofen-1 [28] multisperspectral data. The architecture incorporates 2D convolutions that process both spatial neighborhoods and spectral bands simultaneously, while its encoder-decoder structure with skip connections enables multi-scale feature capture while preserving fine spatial details, making it particularly effective for semantic segmentation tasks.

More recently, Vision Transformer-based architectures have emerged as a powerful alternative for HSS analysis [36]. These models employ self-attention mechanisms to capture long-range dependencies in both spatial and spectral dimensions. Transformers have been also applied to cloud detection using Landsat 8 [37], Sentinel-2 [38, 39], MODIS [40], and Gaofen-1 [28] multispectral images. Although Vision Transformers have shown promising results in various hyperspectral applications [41], their computational complexity and large data requirements can be challenging for operational scenarios [42].

Atmospheric monitoring of greenhouse gases using hyperspectral remote sensing has significantly advanced with missions such as AVIRIS-NG, EMIT, GHGSat, Carbon Mapper, TROPOMI, and MethaneSAT. These instruments detect and quantify greenhouse gases such as CH_4 by leveraging spectral absorption features, primarily in the 1.6 micron and 2.3 micron methane absorption bands. Conventional hyperspectral imaging techniques, including matched filtering [43], band-pass filtering and differential optical absorption spectroscopy (DOAS; [44]), rely on spectral signatures to enhance detection (e.g. CH_4 plumes against background variability [45, 46, 47, 48]). Therefore, these methods often require expert-driven parameter tuning and manual inspection to reduce false positives [6, 49].

Cloud and shadow segmentation in hyperspectral imagery presents a different set of challenges from CH_4 detection, which benefits from well-characterized spectral absorption features. These atmospheric features exhibit high spectral variability across illumination conditions, surface reflectance, and sensor factors, complicating detection efforts [50, 24, 51]. Cloud shadow detection is particularly difficult because dark surfaces like water bodies resemble shadows spectrally, shadowed areas contain diverse land cover types with broad spectral ranges, and thin clouds with high transmittance create shadows that mix with clear pixel characteristics [52]. Consequently, conventional spectral-based classification methods struggle to generalize across diverse scenes [53]. Influence of cloud shadows in the measured top-of-atmosphere radiances can lead to significant biases in atmospehric methane retrievals in turn affecting emission quantification, and therefore correcting for such artifacts is essential for accurately characterizing emission patterns.

3 Data

This study utilizes calibrated and georeferenced Level 1B (L1B) hyperspectral data from MethaneAIR and MethaneSAT [12], with cloud labels derived from Level2 (L2) derived quantities[1] [54]. The L2 products consist of the retrieved fitted parameters obtained from the best fit of a radiative transfer model to the measured L1B spectra. The main products from the CH_4 spectrometer are the retrieved CO_2 and CH_4 vertical column densities (VCDs), and the main product from the O_2 spectrometer is the retrieved surface pressure.

MethaneSAT is a satellite mission developed by MethaneSAT LLC, which is a wholly-owned subsidiary of Environmental Defense Fund. Both MethaneAIR and MethaneSAT data products are available in the public domain.

3.1 MethaneAIR

MethaneAIR [54, 55, 56, 57] is an airborne simulator for the MethaneSAT satellite mission, designed to capture high-resolution spectral imagery for atmospheric composition analysis. MethaneAIR flights are conducted at typically 40,000 ft above ground. MethaneAIR measurements have been conducted on both NCAR GV and Learjet-35 platforms, and over 100 flight hours have been completed across the US, observing major oil and gas production areas. Each individual MethaneAIR campaign maps roughly 100 km x 100 km regions in over 2 hours, which are resolved at high precision and high spatial resolution.

The L1B dataset comprises 508 hyperspectral cubes, each consisting of 1024 spectral bins. MethaneAIR L1B spectra are aggregated by a factor 5 in the across-track direction. We center cropped our images, leading to ~300×178 spatial soundings (along-track × across-track). The spectral range covers key absorption features for CH_4 and CO_2 (1592-1678 nm). In order to reduce computational costs of processing, the O_2 band L1B dataset was not used in this work.

Accompanying each hyperspectral image is a corresponding mask that delineates four distinct categories: clouds, cloud shadows, dark surfaces, and background. These mask labels were created during Level 2 post-processing [54] using a cloud screening algorithm. This algorithm uses the following L2 derived quantities as features in a naive Bayes classifier: the retrieved minus apriori surface pressure (ΔP) from the O_2 band, the O_2-band surface albedo, the ratio of the O_2-band surface albedo with its median, the relative difference between the apriori and the retrieved CO_2, O_2-band H_2O, and CO_2-band H_2O VCDs, and a terrain shadow indicator. Each feature has an associated probability density function (PDF) for clouds, shadows, and terrain shadows obtained from a set of manually labeled MethaneAIR scenes. These PDFs are used to derive cloud, shadow, and "dark surface" probabilities from each L2 feature before combining them with the naive Bayes classifier to yield the final cloud, shadow, and "dark surface" flags. This approach allows cloud flags to be derived even in the absence of O_2-band data and provides some robustness to systematic changes in the L2 products between different flights, mostly caused by temperature variations in the aircraft, that prevent the use of simple thresholding on the L2 fields for cloud screening.

Figure 2 presents a visual representation of the data, showcasing three sample images generated from randomly selected spectral bins. These examples illustrate the diverse spectral information captured across different wavelengths and highlight the spatial resolution of the MethaneAIR instrument.

3.2 MethaneSAT

For MethaneSAT, the CH_4 spectrometer operates between 1598-1683 nm. The satellite does not observe continuously, it instead collects data over a discrete list of targets chosen to cover ~80% of global oil and gas production. Each target is acquired in ~30 seconds and covers a ~220x200 km^2 area with a spatial resolution of ~100x400 m^2 (across-track x along-track, when looking at nadir). Our dataset contains a total of 262 hyperspectral samples. Since MethaneSAT is an agile observing system with more than 20 degree pointing ability, individual scenes can frequently be mapped with wider swaths that are in the vicinity of 400-450 km widths.

[1]To download the data visit `https://dataverse.harvard.edu/dataverse/MAIR_MSAT_CLOUD_SHADOWS`

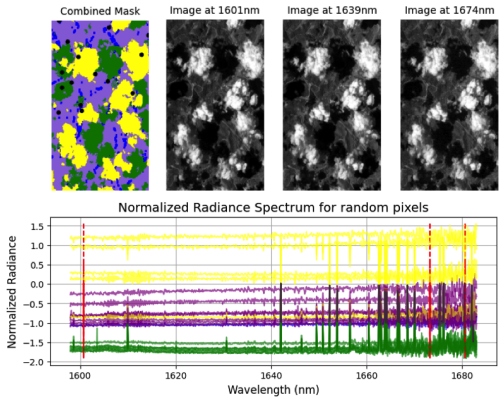


Figure 2: MethaneAIR data example showing classification mask and spectral analysis. The top panel shows a classification mask (purple: background, yellow: clouds, green: shadows, blue: dark surfaces) alongside three images from randomly selected spectral wavelengths. The bottom panel displays normalized radiance spectra from 10 randomly sampled spatial soundings per each class, with colors corresponding to their classification in the mask above. Spectral normalization was performed by computing mean and standard deviation for each spectral band across the entire dataset, then standardizing each spectrum by subtracting the mean and dividing by the standard deviation. This image was captured on September 5, 2023.

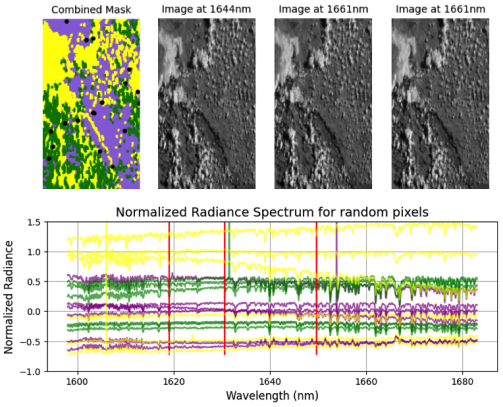


Figure 3: MethaneSAT data example showing classification mask and spectral analysis. The top panel presents the classification mask using the same color scheme as MethaneAIR (purple: background, yellow: clouds, green: shadows) and three representative spectral band images. The bottom panel shows the spectral signatures of 10 randomly selected spatial soundings for each class, with colors indicating their classification status. Spectral normalization was performed by computing mean and standard deviation for each spectral band across the entire dataset, then standardizing each spectrum by subtracting the mean and dividing by the standard deviation. This image was captured on November 18, 2024.

For training our cloud and shadow detection algorithms, we utilize MethaneSAT's cloud screening procedure, which generates masks identifying clouds and shadows in the imagery. Unlike on the aircraft, the satellite instrument is not affected by temperature variability so a simpler cloud screening algorithm can be used effectively. The MethaneSAT cloud screening only uses ΔP, and the background-corrected relative difference between the retrieved and a priori CO_2 VCDs (δCO_2). Cloud flags correspond to ΔP<-20hPa and δCO_2<-2%. Shadow flags correspond to ΔP>20hPa and δCO_2>2%. A potential caveat of the resulting set of flags when using them as a training set is that a pointing error over strong topography can also lead to large ΔP that would be flagged as clouds/shadows. To ensure the high quality of our training set, we discard all images that contain water bodies and full cloudy areas.

Figure 3 presents a visual representation of MethaneSAT's hyperspectral data. The figure displays three sample images from different spectral wavelengths alongside their corresponding classification mask. The bottom panel shows normalized radiance spectra from randomly selected spatial soundings, demonstrating the distinctive spectral signatures captured across the satellite's 220x200 km² footprint. These examples illustrate both the spatial coverage and spectral resolution of the MethaneSAT instrument.

4 Methodology

Our approach for cloud and shadow detection in hyperspectral satellite imagery comprises four main components: data preprocessing, semantic segmentation model development, a training procedure, and a results evaluation procedure. For MethaneSAT, we developed a post-processing component that allows us to evaluate our models despite its variable spatial dimensions. We first establish a preprocessing pipeline that handles missing values and implements strategic normalization steps to ensure consistent input features. Then, we perform a comparison of various semantic segmentation

approaches spanning from traditional machine learning to advanced deep learning architectures. Finally, we detail a training procedure that addresses class imbalance and incorporates data augmentation techniques to enhance model generalization. For MethaneSAT, we present an alternative post-processing strategy that uses fixed-size overlapping patches to accommodate the varying input spatial dimensions.

4.1 Preprocessing Steps

To ensure high-quality input data and optimize model performance, we designed a rigorous preprocessing pipeline. The first step addresses missing values in the hyperspectral data. Any missing values (NaNs) within the dataset are imputed using the mean value along the spectral dimension for each affected pixel, ensuring continuity in the spectral signatures. Missing values may arise, for example, over very dark or very bright features on the landscape.

For spatial standardization, we crop all MethaneAIR L1B files to 300×178 spatial soundings, creating a consistent input size across the dataset. Given the varying input size of MethaneSAT data, we randomly crop patches of 224×224 spatial soundings from the complete images during training, providing a standardized input dimension while increasing effective sample size through data augmentation.

We implement a two-step spectral bin normalization process across all 1024 and 1080 spectral bands of MethaneAIR and MethaneSAT respectively. Specifically, values are clipped to the 1st and 99th percentile to mitigate the impact of extreme outliers that could skew the model training. After clipping, each spectral bin undergoes standardization by subtracting the per-bin mean radiance value and dividing by the per-bin standard deviation. This normalization step ensures that each spectral band contributes equally to the model's learning process.

Before passing the input data to any machine learning model, a batch-wise normalization is performed across all dimensions. In this step, each sample in the batch is normalized by subtracting the mean value calculated across all spectral bands and spatial dimensions (height and width) and dividing by the standard deviation computed across these same dimensions. This approach treats each hyperspectral data cube as a unified entity, normalizing it to have zero mean and unit variance across all its elements. This batch-sample normalization helps reduce the influence of overall intensity variations between different images while maintaining the relative relationships between spectral bands and spatial structures that are crucial to distinguish between clouds, shadows, and clear sky.

4.2 Semantic Segmentation Algorithms

To address the task of cloud and shadow identification in MethaneAIR hyperspectral imagery, we evaluated diverse machine learning methods: Iterative Logistic Regression (LR; [13, 14, 15]), Multilayer Perceptron (MLP; [16, 17, 18, 19, 20]), U-Net [22, 23, 24, 25, 26, 27, 28], and the Spectral Attention Network (SCAN), along with the combined models.

4.2.1 Iterative Logistic Regression (ILR)

We implement the Iterative Logistic Regression (ILR) approach introduced by [14] as a baseline model for our hyperspectral shadow detection system. This established method learns a compact spectral basis that effectively distinguishes between shadowed and non-shadowed spatial soundings through an iterative dimensionality reduction process.

Given an input tensor $X \in \mathbb{R}^{H \times W \times C}$, where H and W are spatial dimensions, and C is the number of spectral channels, the algorithm takes each sounding spectrum $x_i \in \mathbb{R}^C$ in the dataset, compute its mean radiance m_i and apply a normalization transformation:

$$s_i = \log(x_i / m_i)$$

where $\log$ denotes the natural logarithm. This preprocessing step isolates the spectral shape features from overall brightness variations. The core algorithm iteratively extracts spectral components most relevant for cloud and shadow detection:

1. Using log spectra s_i and their associated labels $y_i \in \mathcal{R}^K$, where K is the number of classes, the algorithm trains a logistic regression classifier to find spectral weights w_t that separate each of the K classes.

2. After identifying these discriminative weights, their contribution is projected out from all spectra.

3. This process repeats z times until the classification performance (measured by F1-score) falls below a predefined threshold.

Through this iterative process, we compute a low-dimensional representation with $z = 23$ components (compared to the original 1024 and 1080 spectral bands for MethaneAIR and MethaneSAT respectively), which captures the most salient spectral features for shadow detection.

The learned representation $\beta_i \in \mathcal{R}^z$ is multiplied by a weight matrix $W_{class} \in \mathbb{R}^{z \times K}$ and the Softmax activation function is applied. This way, the probability of belonging to each class can be estimated as:

$$P(y_i|x_i) = \text{Softmax}(W_{class}^T \beta_i)$$

4.2.2 Multilayer Perceptron (MLP)

Building upon the spectral feature extraction approach of ILR, we implement a Multilayer Perceptron (MLP; [58]) model that processes hyperspectral data on a pixel-wise basis. Given an input tensor $X \in \mathbb{R}^{H \times W \times C}$, where H and W are spatial dimensions, and C is the number of spectral channels (1024 for MethaneAIR and 1080 for MethaneSAT), the MLP treats each pixel spectrum $x_i \in \mathbb{R}^C$ independently.

Our MLP implementation processes each pixel's complete spectral signature independently through a neural network architecture optimized for the dimensionality of our data. The network architecture consists of an input layer with C nodes, followed by two hidden layers with 20 nodes each, and an output layer with K nodes corresponding to our target classes. Mathematically, for each pixel spectrum x_i, the probability of belonging to each class can be computed as:

$$\begin{aligned} h_i^{(1)} &= \phi(W^{(1)} x_i + b^{(1)}) \\ h_i^{(2)} &= \phi(W^{(2)} h_i^{(1)} + b^{(2)}) \\ P(y_i|x_i) &= \text{Softmax}(W^{(3)} h_i^{(2)} + b^{(3)}) \end{aligned}$$

where $W^{(1)} \in \mathbb{R}^{20 \times C}$, $W^{(2)} \in \mathbb{R}^{20 \times 20}$, and $W^{(3)} \in \mathbb{R}^{K \times 20}$ are the weight matrices, $b^{(1)} \in \mathbb{R}^{20}$, $b^{(2)} \in \mathbb{R}^{20}$, and $b^{(3)} \in \mathbb{R}^K$ are the bias vectors, and $\phi(\cdot)$ represents the Rectified Linear Unit (ReLU) activation function.

Unlike the ILR approach, which explicitly reduces dimensionality through iterative feature extraction, the MLP learns a direct non-linear mapping from the full spectral signature to class probabilities. This allows the network to capture complex spectral patterns without the need for explicit feature engineering. The compact architecture (only 20 nodes per hidden layer) maintains computational efficiency while providing sufficient capacity to model non-linear relationships between spectral bands that are relevant for cloud and shadow detection.

4.2.3 U-Net Architecture

We adopt the U-Net architecture [22] for hyperspectral semantic segmentation due to its demonstrated effectiveness in dense prediction tasks through its symmetric encoder-decoder structure with skip connections. U-Net has been shown to perform well in hyperspectral segmentation tasks by efficiently capturing both spatial and spectral features across different scales [59, 49].

Given an input hyperspectral image $X \in \mathbb{R}^{B \times C \times H \times W}$, where B is the batch size, C is the number of spectral channels, and H and W are spatial dimensions, our U-Net processes the data through the following components:

Contracting Path (Encoder): The encoder comprises three stages, each consisting of two convolutional layers followed by a downsampling operation. The convolutional layers use 3×3 kernels

with padding to preserve spatial dimensions and a stride of 1. Each convolution is followed by batch normalization and a ReLU activation function. After the convolutions, a 2×2 max pooling operation with stride 2 reduces the spatial resolution by half.

The feature progression for each stage is: Input (C channels) $\rightarrow$ 8 channels $\rightarrow$ 16 channels $\rightarrow$ 32 channels. This systematic increase in feature representation capacity while reducing spatial dimensions enables the network to encode progressively higher-level contextual information at each stage, transitioning from fine-grained spectral patterns to abstract semantic concepts.

This progression increases the feature representation capacity while reducing spatial dimensions, enabling the network to encode higher-level contextual information.

Expanding Path (Decoder): The decoder symmetrically reconstructs the spatial resolution using three upsampling stages. Each stage begins with a transposed convolution (also known as deconvolution) with a kernel size of 3×3, stride 2, and padding to double the spatial resolution. This is followed by the concatenation of the corresponding feature map from the encoder (via skip connections), and two 3×3 convolutions with batch normalization and ReLU activations, similar to the encoder.

The channel progression of each upsampling stage reverses the encoder: 32 channels $\rightarrow$ 16 channels $\rightarrow$ 8 channels $\rightarrow K$ output classes, where each reduction maintains spatial detail while focusing the representation toward final class predictions.

Skip Connections: To retain spatial information lost during downsampling, we employ skip connections that directly pass feature maps from the encoder stages to their corresponding decoder stages. These connections concatenate encoder features with upsampled decoder features along the channel dimension. For example, features from encoder stage 2 (16 channels) are concatenated with upsampled features from decoder stage 2 (16 channels), resulting in a 32-channel input to the following convolutional layers. This fusion of high-resolution spatial detail and abstract semantic information enables accurate boundary delineation and class prediction at the spatial sounding level.

Output Layer: The final layer produces a tensor with dimensions $B \times K \times H \times W$, where K is the number of classes. A softmax function is applied to generate a probability distribution over classes for each pixel:

$$P(y|X) = \text{softmax}(f_{U-Net}(X))$$

This formulation allows the model to produce a dense segmentation mask with per-pixel class labels, suitable for hyperspectral semantic segmentation tasks.

4.2.4 Spectral Channel Attention Network

While U-Net architectures have become standard in semantic segmentation tasks, they often struggle with accurate delineation of cloud and shadow boundaries due to their reliance on fixed receptive fields and limited incorporation of spectral information. Specifically, the transition zones between clouds/shadows and land cover features present challenges where spectral confusion leads to misclassification.

Attention mechanisms have been widely adopted in computer vision, their application to spectral band selection in hyperspectral cloud/shadow segmentation presents special challenges and opportunities. Unlike natural images where spatial attention focuses on semantic objects, spectral attention specifically addresses the band selection problem inherent in hyperspectral analysis. The key insight is that different spectral regions provide varying levels of discrimination between clouds/shadows and surface materials, particularly at boundary regions where spectral mixing occurs.

We reformulate the classical attention mechanism from [60] as a Spectral Channel Attention Network (SCAN), adapting channel-wise attention for hyperspectral band selection in cloud and shadow detection tasks. This approach directly addresses the boundary delineation problem by dynamically weighting the importance of different spectral bands based on their discriminative power for cloud/shadow detection. Rather than treating all spectral bands equally as in conventional approaches, our method learns to emphasize wavelengths that are most informative for distinguishing clouds/shadows from underlying surfaces, particularly in spectrally ambiguous transition zones. The model consists of two main components: a spectral attention module and a pixel-wise classification framework.

Spectral Attention Module: Given an input hyperspectral image $X \in \mathbb{R}^{B\times H\times W\times C}$, where B is the batch size, H and W are spatial dimensions, and C is the number of spectral channels, we apply a channel-wise attention mechanism that learns band-specific importance weights. The attention weights α are computed as:

$$\alpha = \sigma(W_2\text{ReLU}(W_1\bar{x}))$$

where $\bar{x} \in \mathbb{R}^C$ is the spatially averaged input, $W_1 \in \mathbb{R}^{C/16\times C}$ and $W_2 \in \mathbb{R}^{C\times C/16}$ are learnable parameters, and σ is the sigmoid activation function. The attended features are obtained through:

$$X_{att} = X \odot \alpha$$

where $\odot$ represents channel-wise multiplication broadcasting the attention weights across all spatial locations. Intuitively, $X_{att} \in \mathbb{R}^{B\times H\times W\times C}$ has the same input dimension and represents a spectrally-weighted version of the input that emphasizes the most discriminative bands for cloud/shadow detection while suppressing less informative spectral regions.

Classification Framework: The attended features are passed through a fully-connected neural network as follows:

$$f_{MLP}(X) = W_n(\phi(W_{n-1}(...\phi(W_1X_{att})))),$$

and the final probabilities are obtained through a softmax activation function:

$$P(y|X) = \text{Softmax}(f_{MLP}(X_{att})).$$

4.2.5 Combined Models

To leverage the complementary strengths of spatial and spectral approaches, we developed two ensemble methods that combine the U-Net and Spectral Channel Attention Network (SCAN) models: Combined MLP and Combined CNN.

Combined MLP Architecture:

This model fuses U-Net and SCAN predictions using a multilayer perceptron (MLP). The architecture employs two pre-trained base models (U-Net and SCAN) with frozen weights to maintain their individual predictive capabilities. Given the input hyperspectral image $X \in \mathbb{R}^{B\times H\times W\times C}$, the predictions from the base models are:

$$P_{U-Net}(y|X) \in \mathbb{R}^{B\times H\times W\times K}$$
$$P_{SCAN}(y|X) \in \mathbb{R}^{B\times H\times W\times K}$$

where K is the number of classes. These predictions are concatenated channel-wise to form the input to the merging MLP:

$$P_{combined} = \text{Concat}[P_{U-Net}, P_{SCAN}] \in \mathbb{R}^{B\times H\times W\times 2K}$$

The MLP fusion network with M hidden layers processes each pixel independently:

$$f_{fusion}(p) = W_{M+1}(\phi(BN_M(W_M(...\phi(BN_1(W_1P_{combined})))))),$$

where W_i are learnable weight matrices, BN_i denotes batch normalization, and ϕ is the ReLU activation function. Dropout with rate $\delta = 0.2$ is applied after each hidden layer for regularization. The final output provides the fused class probabilities as follows:

$$P(y|X) = \text{Softmax}(f_{fusion}(P_{combined})) \in \mathbb{R}^K$$

Combined CNN Architecture:

This approach employs convolutional layers to merge predictions, preserving spatial context during fusion. Using the same pre-trained U-Net and SCAN models with frozen weights, the concatenated predictions $P_{combined} \in \mathbb{R}^{B\times H\times W\times 2K}$ are processed through a series of convolutional layers:

$$F_1 = \phi(BN_1(\text{Conv}_1(P_{combined}))) \in \mathbb{R}^{B\times H\times W\times C_1}$$

$$F_i = \phi(BN_i(\text{Conv}_i(F_{i-1}))) \in \mathbb{R}^{B\times H\times W\times C_i}, \quad i \in \{2, ..., N-1\}$$

$$F_N = \text{Conv}_N(F_{N-1}) \in \mathbb{R}^{B\times H\times W\times K}$$

where Conv_i represents a 2D convolutional layer with kernel size 3×3 and padding size 1 (except for the final layer which uses 1×1 convolution), preserving original spatial dimensions. BN_i denotes batch normalization, ϕ is the ReLU activation function, and C_i are the channel dimensions of the intermediate feature maps (e.g., $C_1 = 64$, $C_2 = 32$, $C_3 = 16$ in our work). Dropout with rate $\delta = 0.2$ is applied after each intermediate convolutional layer. The final class probabilities are obtained through:

$$P(y|X) = \text{Softmax}(F_N)$$

Unlike the MLP approach, the CNN merger maintains the spatial structure of predictions throughout the fusion process. Each convolutional layer captures local relationships between the predictions of both models, allowing the network to learn spatial patterns in prediction agreement or disagreement. This spatial-aware fusion can be particularly beneficial for complex scenes where the performance of individual models varies spatially.

4.3 Training

We train all models using a weighted cross-entropy loss function that accounts for class imbalance in our dataset. For a batch of N samples, the loss function is defined as:

$$\mathcal{L} = -\sum_{n=1}^{N}\sum_{c=1}^{C} w_c y_{n,c} \log(\hat{y}_{n,c})$$

where C represents the number of classes, w_c denotes the weight assigned to class c, $y_{n,c}$ is the ground truth binary indicator (1 if sample n belongs to class c, and 0 otherwise), and $\hat{y}_{n,c}$ represents the model's predicted probability that sample n belongs to class c. The class weights w_c are computed as the inverse of the class frequencies in the training set, helping to address the inherent imbalance between cloud, shadow, and clear sky spatial soundings.

For optimization, we employ the Adam optimizer with a learning rate determined through cross-validation experiments detailed in Appendix A. To enhance model generalization, we implement data augmentation techniques including random horizontal and vertical flips, and rotations at multiples of 90 degrees. These augmentations help the models learn invariance to common geometric transformations while preserving the physical meaning of the spectral signatures. Training proceeds over a maximum of 100 epochs with randomly shuffled batches of size 32. We employ an early stopping strategy to prevent overfitting, where training is terminated if the validation loss does not improve for 20 consecutive epochs (patience period), and we save the best-performing model checkpoint based on validation performance for final evaluation.

4.4 Post-processing and Evaluation

4.4.1 Post-processing for MethaneSAT

To accommodate the varying input dimensions of MethaneSAT hyperspectral imagery, we implemented a patch-based evaluation strategy. This approach enables consistent model application across diverse acquisition scenarios while maintaining spatial context. Specifically, each variable-sized hyperspectral image is systematically partitioned into overlapping patches of 224×224 spatial soundings with a stride of 112 spatial soundings, ensuring 50% overlap between adjacent regions. The segmentation model, trained exclusively on fixed-dimension patches, processes each segment independently. Subsequently, we employ a weighted averaging scheme in overlapping regions, where model predictions from multiple patches contribute proportionally to the final segmentation map. We show that this simple methodology effectively addresses the dimensional heterogeneity in MethaneSAT data.

4.4.2 Evaluation

To comprehensively evaluate the performance of our cloud and shadow detection models, we employed four key metrics:

- **Accuracy**: The overall correctness of the model, calculated as the proportion of correctly classified spatial soundings across all classes. Although informative, accuracy alone can be misleading for imbalanced datasets where one class significantly outnumbers others (e.g., when clear land spatial soundings vastly outnumber cloud and shadow spatial soundings in the imagery).
- **Precision**: Represents the model's exactness in identifying clouds and shadows, calculated as the ratio of correctly predicted cloud/shadow spatial soundings to all spatial soundings predicted as cloud/shadow by the model.. High precision indicates a low false positive rate, meaning when the model predicts a pixel as cloud or shadow, it is likely to be correct. Precision = $\frac{\text{True Positives}}{\text{True Positives + False Positives}}$
- **Recall**: Measures the completeness of the model in detecting actual clouds and shadows, calculated as the ratio of correctly predicted cloud/shadow pixels to all ground truth cloud/shadow pixels in the dataset. High recall indicates the model's ability to identify most of the actual clouds and shadows. Recall = $\frac{\text{True Positives}}{\text{True Positives + False Negatives}}$
- **F1-Score**: The harmonic mean of precision and recall, providing a balanced measure of the model's performance that accounts for both false positives and false negatives. It is particularly useful when you want to seek a balance between precision and recall and when there is an uneven class distribution. F1-Score = $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}}$.

We use macro-averaging, computing the precision, recall, and F1-score (i.e. we compute the metrics for each class independently and average them to report final results). The reported values include the test set mean performance and standard deviation in 3 cross-validated folds, ensuring a robust evaluation of model performance. See Appendix A for details of the hyperparameter selection.

5 Results

5.1 MethaneAIR

We evaluated six different architectures for cloud and shadow segmentation: Iterative Logistic Regression (ILR), Multi-Layer Perceptron (MLP), U-Net, the Spectral Attention Network (SCAN), and two ensemble approaches—Combined MLP and Combined CNN. Table 1 presents the quantitative comparison of the performance metrics of these models.

Model	Accuracy	F1	Precision	Recall
ILR	73.81+-4.05	62.07+-0.86	61.33+-0.67	72.59+-1.46
MLP	82.49±2.24	71.29±1.02	68.24±1.04	81.42±0.85
U-Net	88.26±0.45	76.24±1.90	72.59±2.13	83.65±1.03
SCAN	86.51±2.90	74.96±0.96	72.17±1.60	83.46±3.13
Combined MLP	88.92±1.80	76.99±6.78	72.79±6.38	86.34±6.32
Combined CNN	**89.42±1.20**	**78.50±3.08**	**74.44±1.89**	**88.97±2.77**

Table 1: Performance comparison of different models for the MethaneAIR dataset.

The Combined CNN architecture achieved the best overall performance with an accuracy of 89.42±1.20% and an F1-score of 78.50±3.08%. Figure 4 illustrates predictions for 9 random samples from our dataset. As can be seen, the model successfully captures complex cloud formations and their associated shadows across diverse landscapes. The model demonstrates particularly strong performance in identifying cloud patterns (yellow regions) with high precision, as evidenced in the second row of samples where the predicted cloud boundaries closely align with the ground truth masks. Shadow detection (green regions) also shows high accuracy, particularly visible in the middle row images where the model correctly identifies the areas affected by cloud shadows despite varying surface reflectance properties underneath. While not achieving the highest overall metrics, the U-Net architecture emerges as the second best-performing model due to its remarkable capacity for detecting

dark surfaces. The performance of this model can be visualized in Figure 5, which demonstrates its effectiveness in challenging detection scenarios.

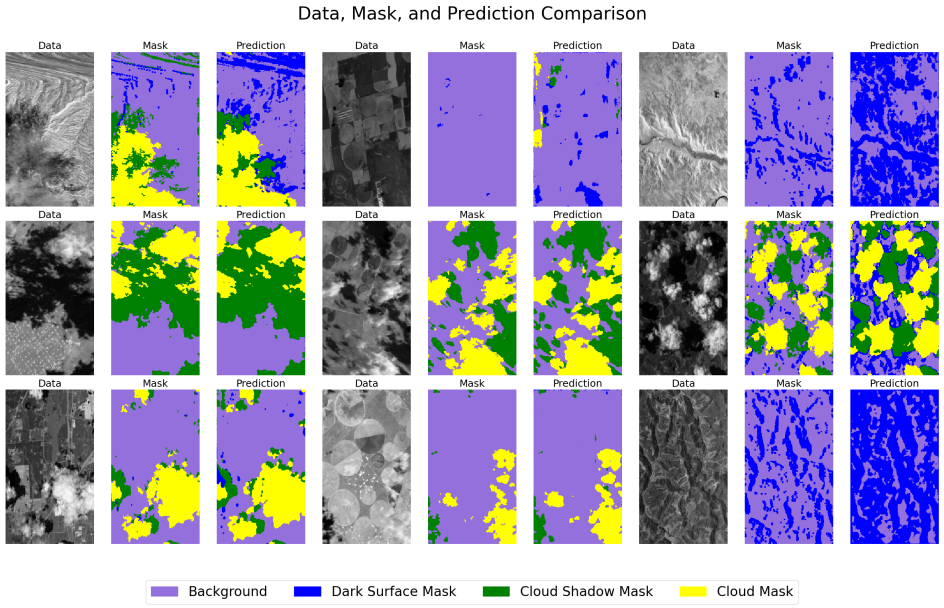


Figure 4: Radiance at 1592nm, labels and predictions of the Combined CNN model. All figures correspond to the test set from the first cross-validation fold (where the dataset was split into 3 parts, with one part held out for testing)

.

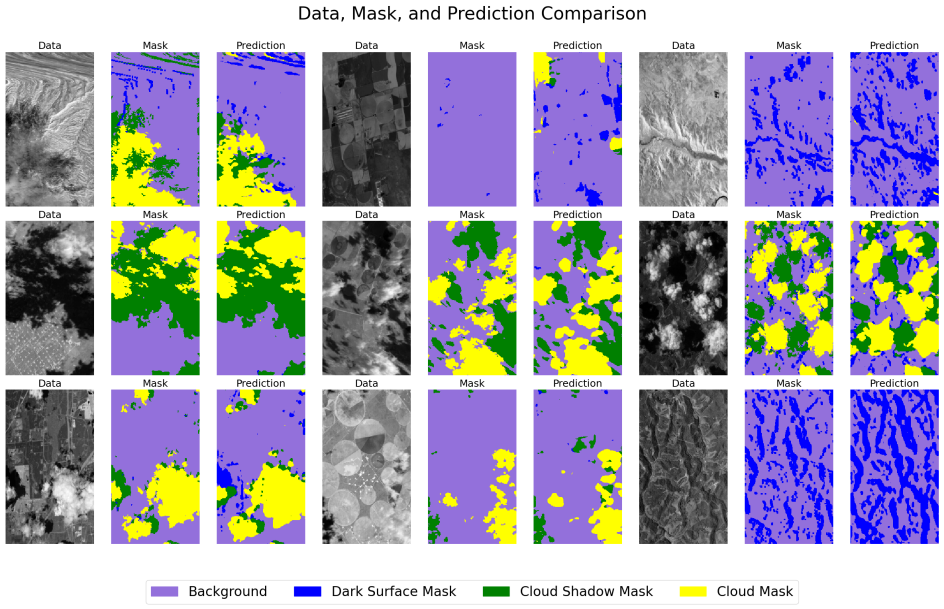
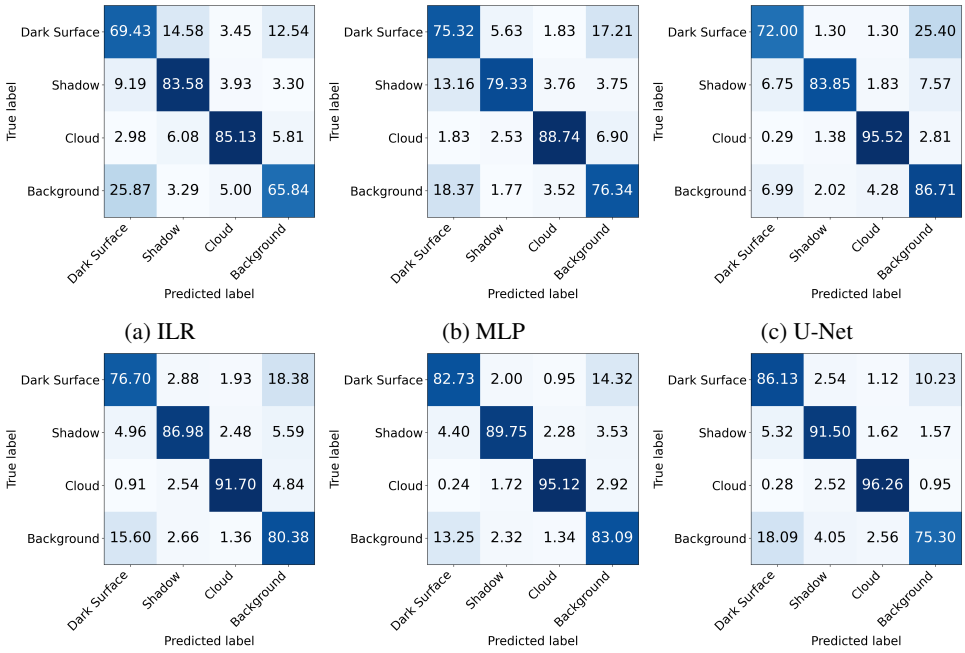


Figure 5: Radiance at 1592nm, ground truth labels, and predictions of the U-Net model for MethaneAIR data. The model demonstrates significantly improved spatial coherence but exhibits over-smoothed boundaries.

Figure 6 presents a comparative view of confusion matrices for all evaluated models on the MethaneAIR dataset. This side-by-side comparison reveals a clear progression in classification performance across model architectures. The ILR model shows significant confusion between background and dark surface classes (18.09% misclassification), while the MLP improves but still



(a) ILR (b) MLP (c) U-Net

(d) SCAN (e) Combined MLP (f) Combined CNN

Figure 6: Confusion matrices for all evaluated models on MethaneAIR data. All the results were computed using the test set over the first cross-validated fold.

struggles with shadow detection. The SCAN model demonstrate substantially higher diagonal values, indicating improved class-specific accuracy, particularly for cloud detection (91.20%). However, it still exhibits confusion between shadows and dark surfaces. The U-Net model achieves best accuracy for background detection (86.71%), and notably reduces the dark surface's false positives. This can be seen in the first row of Figure 6. The Combined MLP approach reduces cloud and shadow misclassifications, while the Combined CNN achieves the highest performance across all classes, with notably improved shadow detection (91.50% accuracy) and significantly reduced confusion between spectrally similar categories.

Figure 7 reveals the distinctive characteristics of each approach across four representative scenes. The ILR and MLP models produce noisy, fragmented predictions due to their pixel-wise processing, particularly evident in the first row where terrain features are inconsistently classified with scattered dark surfaces. In contrast, the SCAN model shows notable improvement in capturing accurate boundaries of clouds and shadows through its spectral attention mechanism, though it still exhibits some noise in spectrally ambiguous regions. Interestingly, while U-Net does not achieve the highest quantitative metrics, visual inspection shows it produces remarkably balanced predictions with significantly less overprediction of dark surfaces compared to other models. This suggests that its lower classification metrics may be attributed to its tendency to generate overly smoothed boundaries around cloud and shadow regions rather than to fundamental misclassifications.

The choice of optimal method should ultimately be determined by domain experts who understand the specific application requirements and downstream processing needs. For instance, in applications where subsequent processing steps include boundary smoothing or where 3-D cloud effects not captured in training masks lead to edge artifacts, U-Net's tendency toward smoother boundaries may actually be advantageous rather than detrimental. Oversmoothing is not necessarily problematic in scenarios where precise edge delineation is less critical than avoiding false positive detections, or where post-processing workflows are designed to handle boundary refinement. Therefore, the "best" model depends on whether the application prioritizes sharp boundary accuracy.

Detailed hyperparameter experiments can be found in Appendix A, and comprehensive predictions for all models can be found in Appendix B.

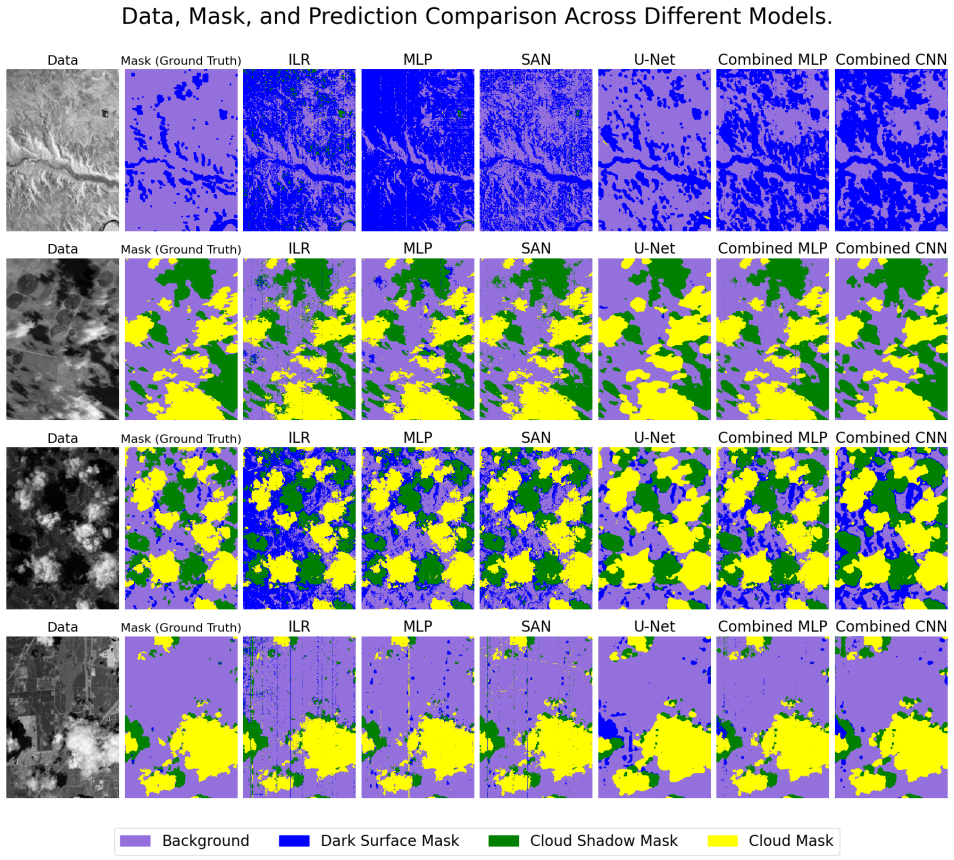


Figure 7: Prediction comparison across all evaluated models for MethaneAIR scenes. All images belongs to the test set and were computed on the first cross-validated fold.

5.2 MethaneSAT

For the MethaneSAT dataset, we observed similar performance trends across models as with MethaneAIR, though with some notable differences in the relative performance of spectral versus spatial approaches. Table 2 presents the performance metrics for each model, again showing the Combined CNN approach achieving the highest scores across all metrics, with an accuracy of 81.96±1.45% and F1-score of 78.80±1.28%.

Model	Accuracy	F1	Precision	Recall
ILR	71.82±4.02	64.35±3.56	70.25±2.57	65.68±1.98
MLP	74.03±3.72	67.11±2.06	69.54±2.86	68.79±0.97
U-Net	78.73±3.23	68.56±0.36	67.87±0.26	71.90±1.76
SCAN	80.33±3.43	71.53±0.75	70.53±0.11	74.73±0.95
Combined MLP	81.32±1.28	78.10±1.72	78.30±1.02	80.35±1.49
Combined CNN	**81.96±1.45**	**78.80±1.28**	**78.85±0.86**	**81.09±1.23**

Table 2: Performance metrics comparison across different models for MethaneSAT data.

Interestingly, for the MethaneSAT data, the SCAN model (accuracy: 80.33±3.43%, F1-score: 71.53±0.75%) outperformed the U-Net model (accuracy: 78.73±3.23%, F1-score: 68.56±0.36%), suggesting that spectral attention mechanisms may be particularly valuable for the unique spectral characteristics of MethaneSAT observations. Figure 8 shows representative examples of the Combined CNN model predictions across varying scenes, demonstrating its effectiveness in accurately identifying clouds and shadows across diverse terrain and atmospheric conditions.

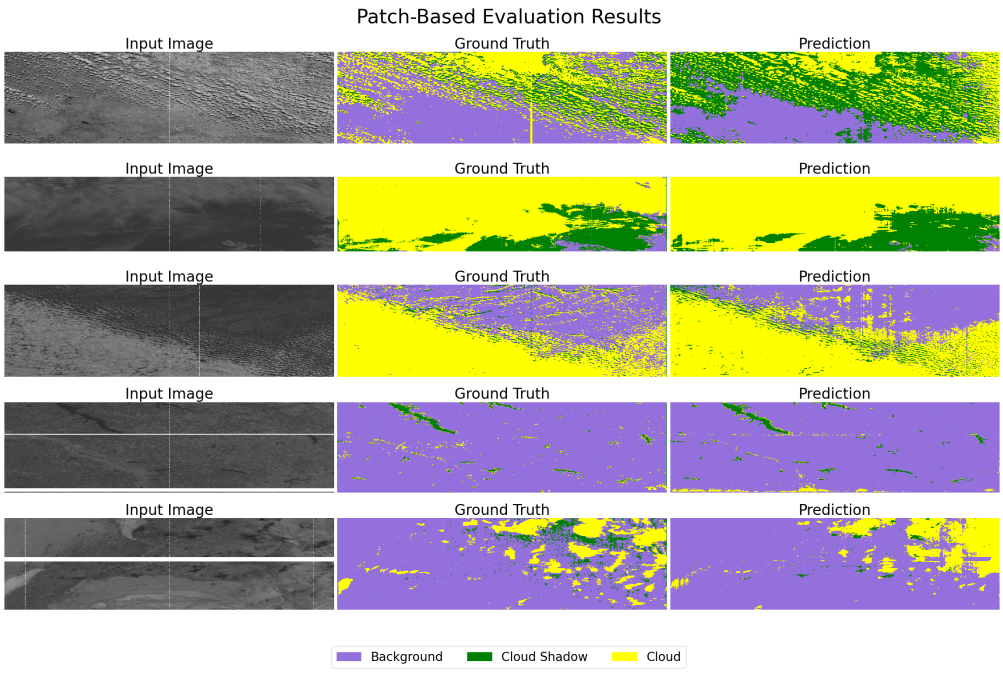
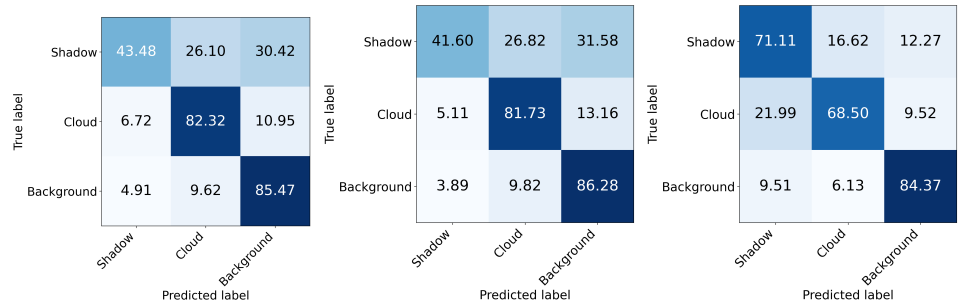


Figure 8: Examples of Combined CNN model predictions on MethaneSAT data showing input images (left), ground truth masks (middle), and model predictions (right) across diverse scenes. Results were computed over the test set using the first cross-validated fold.



(a) ILR (b) MLP (c) U-Net

(d) SCAN (e) Combined MLP (f) Combined CNN

Figure 9: Confusion matrices for all evaluated models on MethaneSAT data. All matrices show results over the test set using the first cross-validated fold.

The confusion matrices for all models on the MethaneSAT dataset (Figure 9) reveal distinctive classification challenges compared to MethaneAIR. Notably, all models demonstrate increased confusion between cloud and shadow classes, with even the Combined CNN showing 12.59%

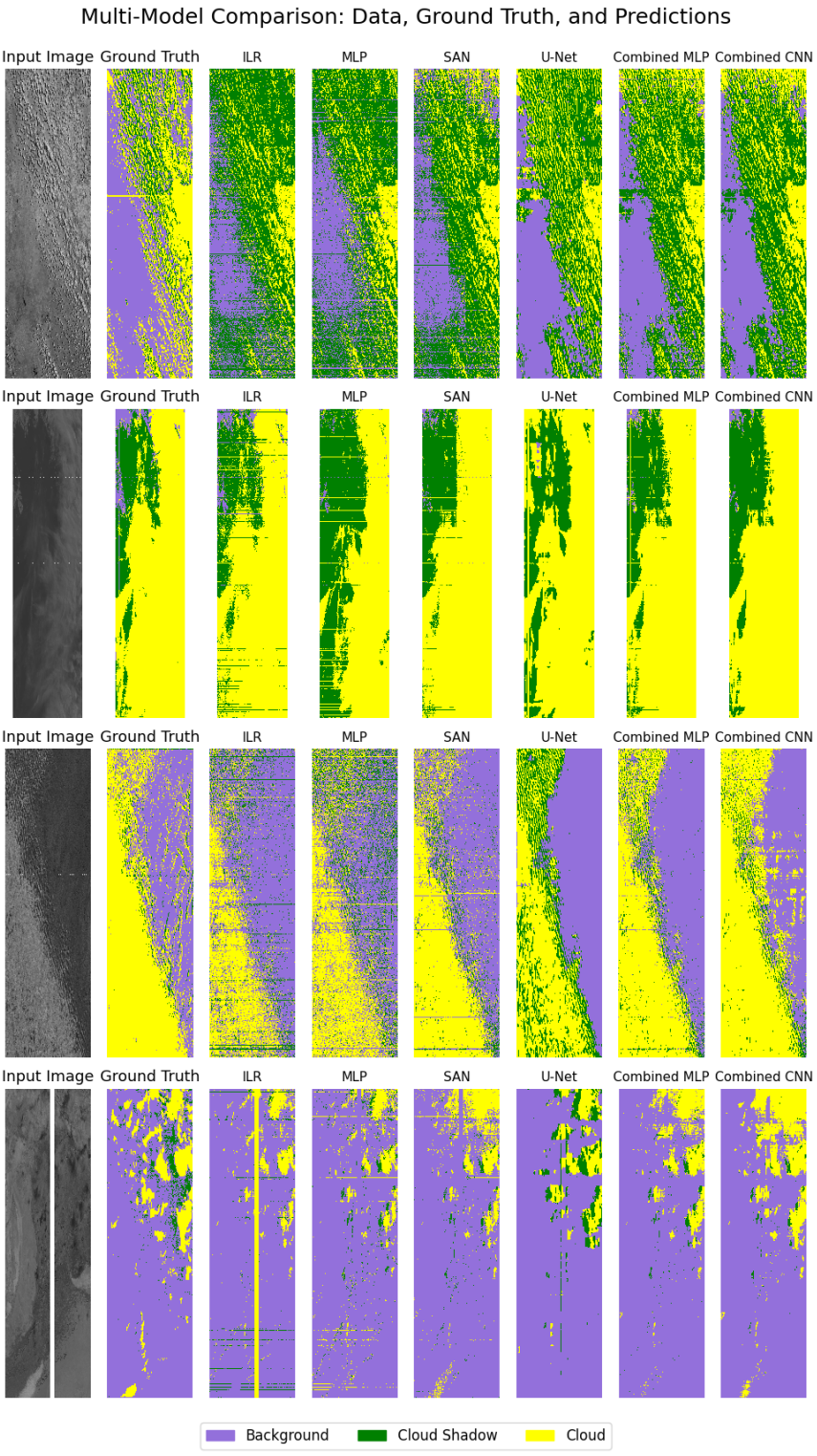


Figure 10: Prediction comparison across all evaluated models for representative MethaneSAT scenes. All the scenes were transposed for better visualization. Displayed images belong to the test set, and predictions were computed using a model trained with the first cross validated fold.

misclassification of cloud spatial soundings as shadows. This pattern suggests greater spectral similarity between these features in the MethaneSAT data. The SCAN model (shadow accuracy: 81.62%) outperforms the U-Net (shadow accuracy: 71.11%) in shadow detection, reversing the trend observed in MethaneAIR data. This reversal highlights the value of spectral attention mechanisms for MethaneSAT spectral characteristics. The Combined MLP substantially improves cloud-shadow discrimination, while the Combined CNN achieves the highest overall performance with class-specific accuracies of 85.02% for background, 79.83% for clouds, and 83.15% for shadows.

The cross-model comparison in Figure 10 further validates the findings from the MethaneAIR analysis. The MLP model produces noisy, fragmented predictions with horizontal striping artifacts particularly visible in the third and fourth rows. The SCAN model demonstrates improved boundary detection but still exhibits noise in spectrally complex regions. The U-Net shows smoother predictions but as can be seen in the first and third row of Figure 10, it tends to overestimate shadows and create more smooth boundaries around the clouds. Both combined models show enhanced performance, with the Combined CNN approach yielding the most balanced results, preserving cloud structure detail while maintaining spatial coherence in shadow regions.

Similar to MethaneAIR, the selection of the most appropriate method for MethaneSAT data should be guided by the specific operational requirements and downstream atmospheric retrieval algorithms. Given MethaneSAT's role in quantitative methane monitoring, applications may benefit from U-Net's conservative approach to boundary delineation, particularly when subsequent atmospheric correction processes can accommodate smoothed cloud edges or when avoiding false cloud detections is more critical than capturing precise cloud boundaries.

5.3 Computational Performance Analysis

To provide a comprehensive understanding of the practical considerations for deploying these models in operational settings, we conducted a detailed evaluation of computational efficiency across all architectures. This analysis is particularly relevant for the MethaneSAT mission, which requires near real-time processing capabilities to support timely methane monitoring.

5.3.1 Experimental Setup

We conducted computational performance measurements on a workstation equipped with an NVIDIA RTX A6000 GPU (48GB VRAM) running CUDA 11.7. All experiments utilized PyTorch 1.12.0 with CUDA acceleration. For each model, we measured:

- Parameters count (millions): Number of parameters to be fitted during training.
- Memory consumption during inference (MB): Memory consumption was evaluated by tracking peak GPU memory allocation during both training and inference processes using PyTorch's native memory profiling utilities. We recorded the maximum memory footprint across complete forward and backward passes to capture the full computational requirements.
- Training time per epoch (seconds): Training performance was assessed by measuring the time required to complete one full epoch across the MethaneSAT training dataset with consistent batch size (8 samples) and optimization parameters. These measurements were averaged over 10 complete training epochs to account for system variability. All models utilized identical loss functions and optimizer configurations.
- Inference time per area (ms/1,000 km²): For inference time measurements, each model processed identical input samples with dimensions matching the MethaneSAT dataset (samples with spatial resolution of 224×224, across 1080 spectral bands). We executed 100 consecutive forward passes for each model after 10 initial warm-up iterations to eliminate cold-start variability. Synchronization barriers were implemented between measurements to ensure accurate timing of GPU operations. Given the spatial resolution of $\sim 100 \times 400$ m^2 (across-track × along-track), each 224×224 sample covers approximately 2,007 km^2, allowing conversion from per-image timing to per-area metrics.

Measurements were averaged over 100 runs for inference and 10 complete training epochs on the MethaneSAT dataset to ensure statistical robustness.

5.3.2 Performance Results

Table 3 presents the computational performance metrics across all evaluated models.

Model	Parameters (M)	Training Time (s/epoch)	Inference Time (ms/1,000 km²)
MLP	0.022	245.4 ± 1.0	1.2 ± 0.0
U-Net	0.113	255.8 ± 7.2	2.1 ± 0.0
SCAN	0.168	290.7 ± 24.2	1.7 ± 0.0
Combined MLP	0.035	296.7 ± 12.4	4.2 ± 0.1
Combined CNN	0.026	326.6 ± 10.3	4.1 ± 0.0

Table 3: Computational Performance of the MLP Model

Our analysis reveals significant differences in computational efficiency across the evaluated models. The lightest model, MLP, demonstrates the fastest inference time (1.2 ms per 1,000 km²), making it well-suited for resource-constrained environments. However, this computational efficiency comes at the cost of significantly lower segmentation accuracy, as demonstrated in our performance evaluation.

The U-Net and SCAN models represent an effective middle ground, with moderate parameter counts (0.113M and 0.168M, respectively) and efficient inference times (2.1 ms and 1.7 ms per 1,000 km²). Despite having more parameters, the SCAN model achieves faster inference, likely due to its efficient spectral attention mechanism that reduces the computational overhead compared to U-Net's multi-scale convolutional operations.

The combined models offer an interesting trade-off. While they require approximately twice the inference time of their individual components (4.2 ms for Combined MLP and 4.1 ms per 1,000 km² for Combined CNN), their parameter counts and memory consumption do not scale linearly with the addition of two base models. This efficiency is achieved through parameter freezing of the base models during training, which allows the fusion components to remain lightweight. Notably, the Combined CNN is marginally faster than the Combined MLP during inference.

6 Discussion

Our comprehensive evaluation of cloud and shadow detection models for hyperspectral satellite imagery reveals critical insights for methane monitoring missions. The performance analysis across diverse model architectures demonstrates a clear progression of capabilities, from the baseline spectral-only approaches to ensemble methods that effectively combine spatial and spectral information.

The baseline ILR model (accuracy: 73.81±4.05 for MethaneAIR) captures basic spectral features but generates highly fragmented predictions with significant noise. This purely spectral approach, while computationally efficient, fails to leverage the spatial relationships crucial for distinguishing between spectrally similar features like shadows and dark surfaces. The MLP model (accuracy: 82.49±2.24%) shows improved performance but still suffers from noise and fragmentation due to its pixel-wise classification approach, as clearly visible in Figures 7 and 10.

The U-Net architecture (accuracy: 88.26±0.45% for MethaneAIR, 78.73±3.23% for MethaneSAT) represents a substantial advancement, leveraging its encoder-decoder structure with skip connections to produce spatially coherent predictions. While U-Net exhibits a tendency toward smoothed boundaries around cloud formations and their shadows, this characteristic can be advantageous depending on the specific application requirements. On the other hand, our proposed SCAN model (accuracy: 86.51±2.90% for MethaneAIR, 80.33±3.43% for MethaneSAT) demonstrates superior ability in detecting accurate boundaries through its spectral attention mechanism, though it still exhibits some noise in spectrally complex regions.

Notably, the SCAN model outperforms U-Net on the MethaneSAT data despite showing slightly lower performance on MethaneAIR. This reversal suggests that spectral attention mechanisms may be particularly valuable for the specific spectral characteristics of MethaneSAT observations, highlighting the importance of architecture selection based on the specific instrument's spectral resolution and characteristics. This finding has significant implications for the development of specialized algorithms for different hyperspectral platforms.

The combined models represent the most promising approach, effectively addressing the limitations of individual architectures. The Combined CNN model achieves the highest performance across all metrics for both datasets (accuracy: 89.42±1.20% for MethaneAIR, 82.99% for MethaneSAT), demonstrating how preserving spatial context during the fusion process yields predictions that combine the detailed boundary sensitivity of SCAN with the spatial coherence of U-Net. This integrated approach proves particularly effective for challenging scenes with complex cloud formations and varying surface reflectance.

Despite models for MethaneAIR and MethaneSAT being each developed, trained, and tested independently, the performance gap in similar architectures (approximately 6-7% in accuracy across models) underscores the challenges in generalizing algorithms performance between different hyperspectral sensors. This discrepancy likely stems from differences in spatial resolution, spectral characteristics, and the complexity of the scenes captured by each instrument. The reduced performance on MethaneSAT data is particularly evident in the confusion matrices, which show increased misclassification between clouds and shadows (12.59% of cloud spatial soundings misclassified as shadows), reflecting the difficulty in distinguishing these spectrally similar features in certain lighting conditions.

From a computational perspective, our analysis reveals important trade-offs between model performance and efficiency. While the MLP offers the fastest inference time (1.2 ms), its limited accuracy makes it unsuitable for operational use. The Combined CNN, despite requiring approximately twice the inference time of individual models (4.1 ms), delivers substantially improved performance without proportional increases in memory consumption or parameter count. This computational efficiency, achieved through parameter freezing and efficient fusion architectures, makes the Combined CNN viable for operational deployment in satellite missions with reasonable computational constraints.

These findings have direct implications for the MethaneSAT mission, which requires accurate identification of atmospheric artifacts to ensure reliable methane retrievals. The superiority of the Combined CNN approach in handling complex scenes with varying surface reflectance is particularly relevant for global methane monitoring, where diverse terrain and atmospheric conditions can significantly impact detection accuracy. By effectively eliminating clouds and shadows from analysis, this approach can substantially improve the reliability of methane concentration mapping, supporting global efforts to identify and quantify methane emissions.

7 Conclusions

This study addresses a critical challenge in atmospheric methane monitoring: the accurate detection of clouds and shadows in hyperspectral satellite imagery, which is essential for reliable retrieval of methane concentrations, especially when a sensor combines a wide swath, fine spatial resolution, and high precision, as achieved by MethaneSAT and MethaneAIR. These sensors are designed to provide holistic assessments of point sources, area sources, and regional totals, from a single image, and interference by clouds and by cloud and terrain shadows can introduce significant errors. Highly reliable cloud and shadow masking algorithms are critical for next-generation methane monitoring missions that aim to track and mitigate greenhouse gas emissions as part of global climate change efforts.

Our comprehensive evaluation of semantic segmentation models for MethaneSAT and MethaneAIR reveals a clear progression in performance from traditional spectral-based methods (ILR, MLP) to advanced deep learning architectures (U-Net, SCAN), with each approach demonstrating distinct strengths and limitations. While spectral-based methods lack spatial context, U-Net produces spatially coherent but over-smoothed predictions, and SCAN excels at boundary detection but exhibits noise in complex regions. The Combined CNN model emerges as the superior approach, achieving the highest performance metrics for both MethaneAIR (accuracy: 89.42±1.20%, F1-score: 78.50±3.08%) and MethaneSAT data (accuracy: 82.99%, F1-score: 80.45%). By preserving spatial context during fusion, this model successfully integrates the precise boundary detection capabilities of spectral attention mechanisms with the spatial coherence of convolutional architectures. This is the critical element needed to attain performance that spans over wide areas down to individual facilities. The performance differences between datasets highlight the importance of tailoring algorithmic approaches to specific sensor characteristics and use cases, with SCAN notably outperforming U-Net on MethaneSAT data despite showing slightly lower performance on MethaneAIR.

From a computational perspective, our analysis demonstrates that the Combined CNN model offers an effective balance between performance and efficiency. While requiring moderately increased computational resources compared to individual models (4.1 ms inference time versus 1.7-2.1 ms per 1,000 km²), its significantly improved segmentation accuracy justifies this overhead for operational deployment in satellite missions where accurate atmospheric artifact detection is critical for reliable methane retrievals and subsequent emission quantification.

The patch-based evaluation strategy demonstrated for MethaneSAT's variable-sized inputs enhances the practical applicability of these methods in operational contexts. By providing more accurate masks for clouds and shadows, these approaches can significantly improve the reliability of methane concentration mapping, contributing to global efforts to monitor and mitigate methane emissions, a critical component in addressing climate change.

Future research should focus on developing more efficient fusion architectures, and investigating transfer learning techniques to improve the model's performance. The Combined CNN approach represents a significant advancement in hyperspectral image segmentation for atmospheric artifact detection, balancing high performance with reasonable computational requirements for operational satellite-based methane monitoring.

8 Acknowledgments

Funding for MethaneSAT and MethaneAIR activities was provided in part by Anonymous, Arnold Ventures, The Audacious Project, Ballmer Group, Bezos Earth Fund, The Children's Investment Fund Foundation, Heising-Simons Family Fund, King Philanthropies, Robertson Foundation, Skyline Foundation and Valhalla Foundation. For a more complete list of funders, please visit `www.methanesat.org`. We thank the AstroAI and EarthAI institutes at the Center for Astrophyiscs | Harvard & Smithsonian for useful discussions and guidance. CG was supported by AstroAI at the Center for Astrophysics | Harvard and Smithsonian.

References

[1] G. Myhre, D. Shindell, F.-M. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J.-F. Lamarque, D. Lee, B. Mendoza, T. Nakajima, A. Robock, G. Stephens, T. Takemura, and H. Zhang. *Anthropogenic and natural radiative forcing*, pages 659–740. Cambridge University Press, Cambridge, UK, 2013.

[2] Maryam Etminan, Gunnar Myhre, Eleanor J. Highwood, and Keith P. Shine. Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophysical Research Letters*, 43:12,614 – 12,623, 2016.

[3] Drew Shindell, Johan C. I. Kuylenstierna, Elisabetta Vignati, Rita van Dingenen, Markus Amann, Zbigniew Klimont, Susan C. Anenberg, Nicholas Muller, Greet Janssens-Maenhout, Frank Raes, Joel Schwartz, Greg Faluvegi, Luca Pozzoli, Kaarle Kupiainen, Lena Höglund-Isaksson, Lisa Emberson, David Streets, V. Ramanathan, Kevin Hicks, N. T. Kim Oanh, George Milly, Martin Williams, Volodymyr Demkine, and David Fowler. Simultaneously mitigating near-term climate change and improving human health and food security. *Science*, 335(6065):183–189, 2012.

[4] World Economic Forum. Global methane pledge: which countries are cutting emissions? *World Economic Forum*, August 2024. Accessed: 2025-07-17.

[5] COP28 UAE. Oil & gas decarbonization charter. `https://www.ogdc.org/`, 2023. Launched at COP28, Dubai.

[6] Christian Frankenberg, Andrew K. Thorpe, David R. Thompson, Glynn Hulley, Eric Adam Kort, Nick Vance, Jakob Borchardt, Thomas Krings, Konstantin Gerilowski, Colm Sweeney, Stephen Conley, Brian D. Bue, Andrew D. Aubrey, Simon Hook, and Robert O. Green. 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.

[7] D. J. Varon, D. Jervis, J. McKeever, I. Spence, D. Gains, and D. J. Jacob. High-frequency monitoring of anomalous methane point sources with multispectral sentinel-2 satellite observations. *Atmospheric Measurement Techniques*, 14(4):2771–2785, 2021.

[8] Luis Guanter, Itziar Irakulis-Loitxate, Javier Gorroño, Elena Sánchez-García, Daniel H. Cusworth, Daniel J. Varon, Sergio Cogliati, and Roberto Colombo. Mapping methane point emissions with the PRISMA spaceborne imaging spectrometer. *Remote Sensing of Environment*, 265:112671, November 2021.

[9] J.P. Veefkind, I. Aben, K. McMullan, H. Förster, J. de Vries, G. Otter, J. Claas, H.J. Eskes, J.F. de Haan, Q. Kleipool, M. van Weele, O. Hasekamp, R. Hoogeveen, J. Landgraf, R. Snel, P. Tol, P. Ingmann, R. Voors, B. Kruizinga, R. Vink, H. Visser, and P.F. Levelt. Tropomi on the esa sentinel-5 precursor: A gmes mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sensing of Environment*, 120:70–83, 2012. The Sentinel Missions - New Opportunities for Science.

[10] Marc Watine-Guiu, Daniel J. Varon, Itziar Irakulis-Loitxate, Nicholas Balasus, and Daniel J. Jacob. Geostationary satellite observations of extreme and transient methane emissions from oil and gas infrastructure. *Proceedings of the National Academy of Sciences*, 120(52):e2310797120, 2023.

[11] R. R Rohrschneider, S. Wofsy, J. E. Franklin, J. Benmergui, J. C Soto, and Spencer B Davis. The methanesat mission. *Small Satellite Conference*.

[12] E. K. Conway, A. H. Souri, J. Benmergui, K. Sun, X. Liu, C. Staebell, C. Chan Miller, J. Franklin, J. Samra, J. Wilzewski, S. Roche, B. Luo, A. Chulakadabba, M. Sargent, J. Hohl, B. Daube, I. Gordon, K. Chance, and S. Wofsy. Level0 to level1b processor for methaneair. *Atmospheric Measurement Techniques*, 17(4):1347–1362, 2024.

[13] Edisanter Lo and Emmett Ientilucci. Target detection in hyperspectral Imaging using logistic regression. In Miguel Velez-Reyes and David W. Messinger, editors, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXII*, volume 9840, page 98400W. International Society for Optics and Photonics, SPIE, 2016.

[14] Core Francisco Park, Maya Nasr, Manuel Pérez-Carrasco, Eleanor Walker, Douglas Finkbeiner, and Cecilia Garraffo. Hyperspectral shadow removal with iterative logistic regression and latent parametric linear combination of gaussians, 2023.

[15] Rajneesh Kumar Gautam and Sudhir Nadda. Hyperspectral image prediction using logistic regression model. In Arti Noor, Kriti Saroha, Emil Pricop, Abhijit Sen, and Gaurav Trivedi, editors, *Proceedings of Emerging Trends and Technologies on Intelligent Systems*, pages 283–293, Singapore, 2023. Springer Nature Singapore.

[16] P. H. SWAIN J. A. BENEDIKTSSON and O. K. ERSOY. Conjugate-gradient neural networks in classification of multisource and very-high-dimensional remote sensing data. *International Journal of Remote Sensing*, 14(15):2883–2903, 1993.

[17] H. Yang. A back-propagation neural network for mineralogical mapping from aviris data. *International Journal of Remote Sensing*, 20(1):97–110, 1999.

[18] Bin Tian, M.A. Shaikh, M.R. Azimi-Sadjadi, T.H.V. Haar, and D.L. Reinke. A study of cloud classification with neural networks using spectral and textural features. *IEEE Transactions on Neural Networks*, 10(1):138–151, 1999.

[19] Alireza Taravat, Simon Proud, Simone Peronaci, Fabio Del Frate, and Natascha Oppelt. Multilayer perceptron neural networks model for meteosat second generation seviri daytime cloud masking. *Remote Sensing*, 7(2):1529–1539, 2015.

[20] Luis Gómez-Chova, Gonzalo Mateo-García, Jordi Muñoz-Marí, and Gustau Camps-Valls. Cloud detection machine learning algorithms for proba-v. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 2251–2254, 2017.

[21] Zahid Hassan Tushar, Adeleke Ademakinwa, Jianwu Wang, Zhibo Zhang, and Sanjay Purushotham. Cloudunet: Adapting unet for retrieving cloud properties. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 7163–7167, 2024.

[22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[23] Libin Jiao, Lian-Zhi Huo, Changmiao Hu, and Ping Tang. Refined unet: Unet-based refinement network for cloud and shadow precise segmentation. *Remote Sensing*, 12:2001, 06 2020.

[24] Marc Wieland, Yu Li, and Sandro Martinis. Multi-sensor cloud and cloud shadow segmentation with a convolutional neural network. *Remote Sensing of Environment*, 230:111203, 2019.

[25] Shoukuan Miao, Min Xia, Ming Qian, Yonghong Zhang, Jia Liu, and Haifeng Lin and. Cloud/shadow segmentation based on multi-level feature enhanced network for remote sensing imagery. *International Journal of Remote Sensing*, 43(15-16):5940–5960, 2022.

[26] Nicholas Wright, John MA Duncan, J Nik Callow, Sally E Thompson, and Richard J George. Clouds2mask: a novel deep learning approach for improved cloud and cloud shadow masking in sentinel-2 imagery. *Remote Sensing of Environment*, 306:114122, 2024.

[27] Xian Li, Xiaofei Yang, Xutao Li, Shijian Lu, Yunming Ye, and Yifang Ban. Gcdb-unet: A novel robust cloud detection approach for remote sensing images. *Knowledge-Based Systems*, 238:107890, 2022.

[28] Yuhao Tan, Wenhao Zhang, Xiufeng Yang, Qiyue Liu, Xiaofei Mi, Juan Li, Jian Yang, and Xingfa Gu. Cloud and cloud shadow detection of gf-1 images based on the swin-unet method. *Atmosphere*, 14(11), 2023.

[29] Muhammad Ahmad, Sidrah Shabbir, Swalpa Kumar Roy, Danfeng Hong, Xin Wu, Jing Yao, A. Khan, Manuel Mazzara, Salvatore Distefano, and Jocelyn Chanussot. Hyperspectral image classification—traditional to deep models: A survey for future prospects. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:968–999, 2021.

[30] Reaya Grewal, Singara Singh Kasana, and Geeta Kasana. Hyperspectral image segmentation: a comprehensive survey. *Multimedia Tools and Applications*, 82:1–54, 10 2022.

[31] Vinod Kumar, Ravi Shankar Singh, Medara Rambabu, and Yaman Dua. Deep learning for hyperspectral image classification: A survey. *Computer Science Review*, 53:100658, 2024.

[32] M.E. Paoletti, J.M. Haut, J. Plaza, and A. Plaza. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158:279–317, 2019.

[33] Jun Li, José M. Bioucas-Dias, and Antonio Plaza. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4085–4098, 2010.

[34] Kavitha Munishamaiaha, Senthil Kumar Kannan, DhilipKumar Venkatesan, Michał Jasiński, Filip Novak, Radomir Gono, and Zbigniew Leonowicz. Hyperspectral image classification with deep cnn using an enhanced elephant herding optimization for updating hyper-parameters. *Electronics*, 12(5), 2023.

[35] Wei Hu, Yangyu Huang, Wei Li, Fan Zhang, and Hengchao Li. Deep convolutional neural networks for hyperspectral image classification. *J. Sensors*, 2015:258619:1–258619:12, 2015.

[36] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *ArXiv*, abs/2010.11929, 2020.

[37] Jiayi Li and Qunming Wang. Csdformer: A cloud and shadow detection method for landsat images based on transformer. *International Journal of Applied Earth Observation and Geoinformation*, 129:103799, 2024.

[38] Bin Zhang, Yongjun Zhang, Yansheng Li, Yi Wan, and Yongxiang Yao. Cloudvit: A lightweight vision transformer network for remote sensing cloud detection. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023.

[39] Xiangbing Yan, Jia Song, Yangxiaoyue Liu, Shanlong Lu, Yuyue Xu, Chenyan Ma, and Yunqiang Zhu. A transformer-based method to reduce cloud shadow interference in automatic lake water surface extraction from sentinel-2 imagery. *Journal of Hydrology*, 620:129561, 2023.

[40] Wenxuan Ge, Xubing Yang, Rui Jiang, Wei Shao, and Li Zhang. Cd-ctfm: A lightweight cnn-transformer network for remote sensing cloud detection fusing multiscale features. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:4538–4551, 2024.

[41] Danfeng Hong, Zhu Han, Jing Yao, Lianru Gao, Bing Zhang, Antonio Plaza, and Jocelyn Chanussot. Spectralformer: Rethinking hyperspectral image classification with transformers. *CoRR*, abs/2107.02988, 2021.

[42] Muhammad Ahmad, Salvatore Distefano, Adil Mehmood Khan, Manuel Mazzara, Chenyu Li, Hao Li, Jagannath Aryal, Yao Ding, Gemine Vivone, and Danfeng Hong. A comprehensive survey for hyperspectral image classification: The evolution from conventional to transformers and mamba models. *Neurocomputing*, page 130428, 2025.

[43] Daniel R Fuhrmann, Edward J Kelly, and Ramon Nitzberg. A cfar adaptive matched filter detector. *IEEE Trans. Aerosp. Electron. Syst*, 28(1):208–216, 1992.

[44] Ulrich Platt and Jochen Stutz. *Differential Absorption Spectroscopy*, pages 135–174. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[45] D. R. Thompson, I. Leifer, H. Bovensmann, M. Eastwood, M. Fladeland, C. Frankenberg, K. Gerilowski, R. O. Green, S. Kratwurst, T. Krings, B. Luna, and A. K. Thorpe. Real-time remote detection and measurement for airborne imaging spectroscopy: a case study with methane. *Atmospheric Measurement Techniques*, 8(10):4383–4397, 2015.

[46] Markus D. Foote, Philip E. Dennison, Andrew K. Thorpe, David R. Thompson, Siraput Jongaramrungruang, Christian Frankenberg, and Sarang C. Joshi. Fast and accurate retrieval of methane concentration from imaging spectrometer data using sparsity prior. *IEEE Transactions on Geoscience and Remote Sensing*, 58(9):6480–6492, 2020.

[47] Thibaud Ehret, Aurélien De Truchis, Matthieu Mazzolini, Jean-Michel Morel, Alexandre d'Aspremont, Thomas Lauvaux, Riley Duren, Daniel Cusworth, and Gabriele Facciolo. Global tracking and quantification of oil and gas methane emissions from recurrent sentinel-2 imagery. *Environmental Science & Technology*, 56(14):10517–10529, July 2022.

[48] Satish Kumar, Ivan Arevalo, ASM Iftekhar, and B S Manjunath. Methanemapper: Spectral absorption aware hyperspectral transformer for methane detection, 2023.

[49] V Růžička, G Mateo-Garcia, L Gómez-Chova, A Vaughan, L Guanter, and A Markham. Semantic segmentation of methane plumes with hyperspectral machine learning models. *Scientific Reports*, 13(1), 2023.

[50] Zhe Zhu, Shixiong Wang, and Curtis E. Woodcock. Improvement and expansion of the fmask algorithm: cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images. *Remote Sensing of Environment*, 159:269–277, 2015.

[51] Dengfeng Chai, Shawn Newsam, Hankui K. Zhang, Yifan Qiu, and Jingfeng Huang. Cloud and cloud shadow detection in landsat imagery based on deep convolutional neural networks. *Remote Sensing of Environment*, 225:307–316, 2019.

[52] Han Zhai, Hongyan Zhang, Liangpei Zhang, and Pingxiang Li. Cloud/shadow detection based on spectral indices for multi/hyperspectral optical remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144:235–253, 2018.

[53] Zhengxin Wang, Longlong Zhao, Jintao Meng, Yu Han, Xiaoli Li, Ruixia Jiang, Jinsong Chen, and Hongzhong Li. Deep learning-based cloud detection for optical remote sensing images: A survey. *Remote Sensing*, 16(23), 2024.

[54] C. Chan Miller, S. Roche, J. S. Wilzewski, X. Liu, K. Chance, A. H. Souri, E. Conway, B. Luo, J. Samra, J. Hawthorne, K. Sun, C. Staebell, A. Chulakadabba, M. Sargent, J. S. Benmergui, J. E. Franklin, B. C. Daube, Y. Li, J. L. Laughner, B. C. Baier, R. Gautam, M. Omara, and S. C. Wofsy. Methane retrieval from methaneair using the co_2 proxy approach: a demonstration for the upcoming methanesat mission. *Atmospheric Measurement Techniques*, 17(18):5429–5454, 2024.

[55] Apisada Chulakadabba, Maryann Sargent, Thomas Lauvaux, Joshua S Benmergui, Jonathan E Franklin, Christopher Chan Miller, Jonas S Wilzewski, Sébastien Roche, Eamon Conway, Amir H Souri, et al. Methane point source quantification using methaneair: A new airborne imaging spectrometer. *EGUsphere*, 2023:1–22, 2023.

[56] L. Guanter, J. Warren, M. Omara, A. Chulakadabba, J. Roger, M. Sargent, J. E. Franklin, S. C. Wofsy, and R. Gautam. Remote sensing of methane point sources with the methaneair airborne spectrometer. *EGUsphere*, 2025:1–22, 2025.

[57] J. D. Warren, M. Sargent, J. P. Williams, M. Omara, C. C. Miller, S. Roche, K. MacKay, E. Manninen, A. Chulakadabba, A. Himmelberger, J. Benmergui, Z. Zhang, L. Guanter, S. Wofsy, and R. Gautam. Sectoral contributions of high-emitting methane point sources from major u.s. onshore oil and gas producing basins using airborne measurements from methaneair. *EGUsphere*, 2024:1–22, 2024.

[58] Paul Werbos. *Applications of advances in nonlinear sensitivity analysis*, volume 38, pages 762–770. 01 1970.

[59] Siqi Wei, Yafei Liu, Mengshan Li, Haijun Huang, Xin Zheng, and Lixin Guan. Dccaps-unet: A u-shaped hyperspectral semantic segmentation model based on the depthwise separable and conditional convolution capsule network. *Remote Sensing*, 15(12), 2023.

[60] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

A Appendix A: Hyperparameter Selection

For hyperparameter optimization, we employed a systematic approach using a 3-fold cross-validation strategy on a randomly selected 10% subsample of the training data.

For each model, we evaluated performance across learning rates $\{1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 5 \times 10^{-3}, 1 \times 10^{-2}\}$, using 3-fold cross-validation on the training set. Table 4 shows the chosen learning rates for each model architecture for MethaneAIR and MethaneSAT.

Model	**MethaneAIR**	**MethaneSAT**
ILR	1×10^{-2}	1×10^{-2}
MLP	1×50^{-3}	1×10^{-2}
SCAN	1×10^{-3}	1×10^{-3}
U-Net	1×10^{-3}	5×10^{-3}
Combined MLP	1×10^{-2}	5×10^{-4}
Combined CNN	1×10^{-2}	5×10^{-4}

Table 4: Best learning rates for each data source and model architecture.

For all models, we employed the Adam optimizer with default $\beta_1 = 0.9$ and $\beta_2 = 0.999$ parameters. We applied class weighting to address class imbalance, with weights computed as the inverse of the class frequencies in the training set. No specific normalization was applied beyond the preprocessing steps described in Section 4.1.

The remaining hyperparameters were set according to the architectural specifications detailed in the methodology section. For the MLP, we used hidden layer dimensions of [20, 20]. The U-Net architecture maintained the channel dimensions specified in Section 3.2.3, while the SCAN model used a reduction ratio of 16 for the attention mechanism. The Combined MLP employed hidden dimensions of [256, 128] with a dropout rate of 0.2, and the Combined CNN used channel dimensions of [64, 32, 16] with the same dropout rate.

These hyperparameter configurations were consistently applied across all experimental evaluations to ensure fair comparison between the different architectures.

B Appendix B: Detailed Model Predictions

This appendix provides a comprehensive visual comparison of predictions all evaluated models across both the MethaneAIR and MethaneSAT datasets. These supplementary results validate and expand upon the analyses presented in Section 5.

B.1 MethaneAIR Predictions

Figures 11 through 14 present detailed prediction results for all evaluated models on the MethaneAIR dataset. These visualizations offer insight into the specific capabilities and limitations of each approach when applied to identical input data.

The ILR predictions (Figure 11) exhibit significant fragmentation and noise, particularly evident in regions with complex terrain features. This visual noise corresponds to the high misclassification rates shown in the confusion matrix presented in the main results section, confirming the limitations of purely spectral approaches when spatial context is not considered.

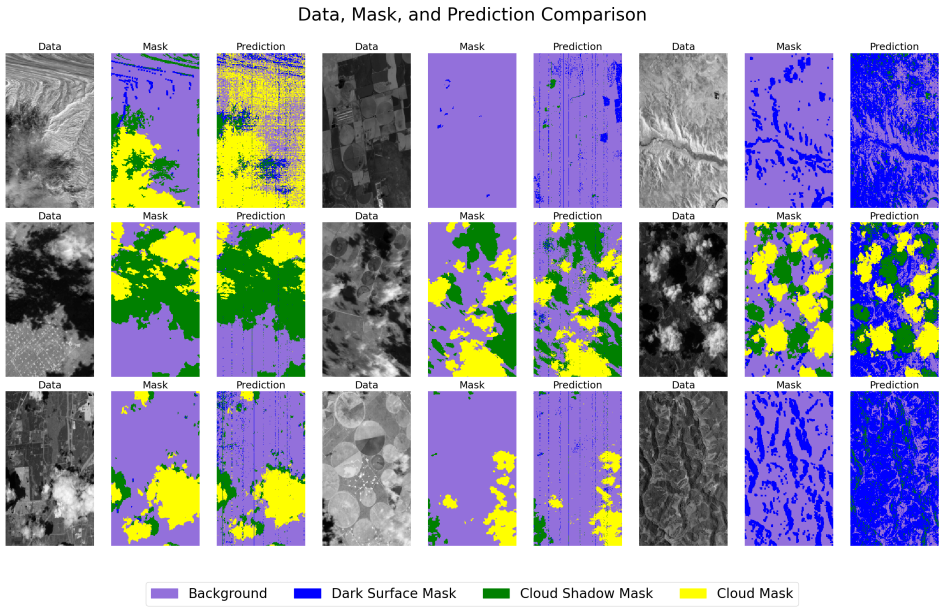


Figure 11: Radiance at 1592nm, ground truth labels, and predictions of the ILR model for MethaneAIR data. The model exhibits significant noise and fragmentation, particularly in regions with complex terrain features.

The MLP model results (Figure 12) show improved coherence compared to ILR but still produce speckled predictions characteristic of pixel-wise classification approaches. While the model better captures the general distribution of clouds and shadows, it struggles with consistent classification, especially in areas with varying surface reflectance.

U-Net predictions (Figure 5) demonstrate a substantial improvement in spatial coherence, with smoother, more contiguous regions for each class. However, as noted in the main results, the model's convolutional architecture tends to produce over-smoothed boundaries that do not precisely capture the intricate edges of cloud formations and their shadows.

The SCAN model results (Figure 13) illustrate its enhanced boundary detection capabilities compared to U-Net, with more precise delineation of cloud and shadow edges. This improved boundary precision comes with some residual noise in spectrally complex regions, representing the trade-off between spatial coherence and spectral fidelity.

The Combined MLP approach (Figure 14) demonstrates the advantages of ensemble learning, with predictions that balance the strengths of individual models. By adaptively weighting the contributions of both U-Net and SCAN, this fusion approach reduces both the over-smoothing tendency of U-Net and the noise artifacts present in SCAN predictions.

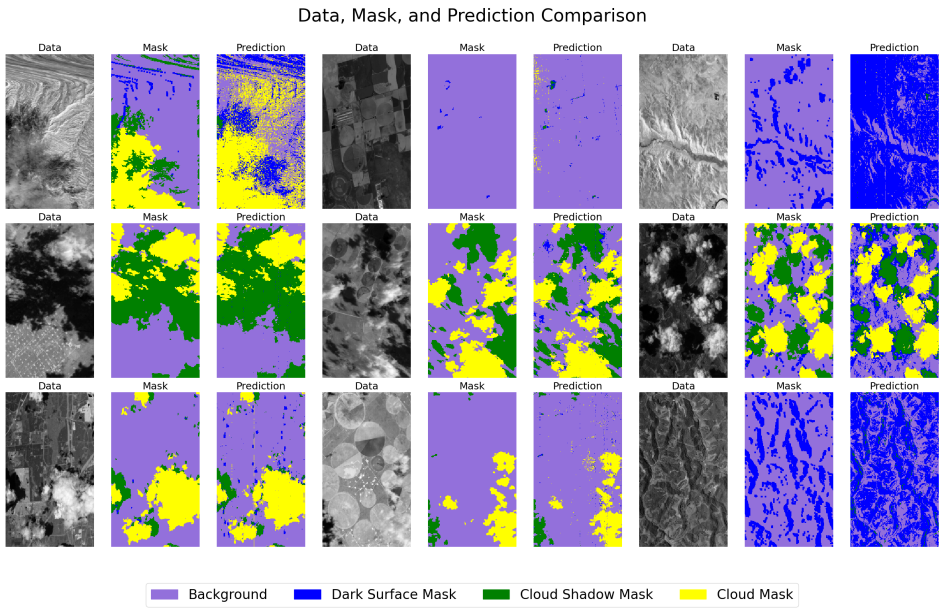


Figure 12: Radiance at 1592nm, ground truth labels, and predictions of the MLP model for MethaneAIR data. While showing improved coherence compared to ILR, the model still produces speckled predictions.

Finally, the Combined CNN results (Figure 4) achieve the most balanced and accurate predictions across all scenes. The model successfully integrates the precise boundary detection of SCAN with the spatial coherence of U-Net, validating the quantitative superiority demonstrated in the confusion matrices presented in the main results section.

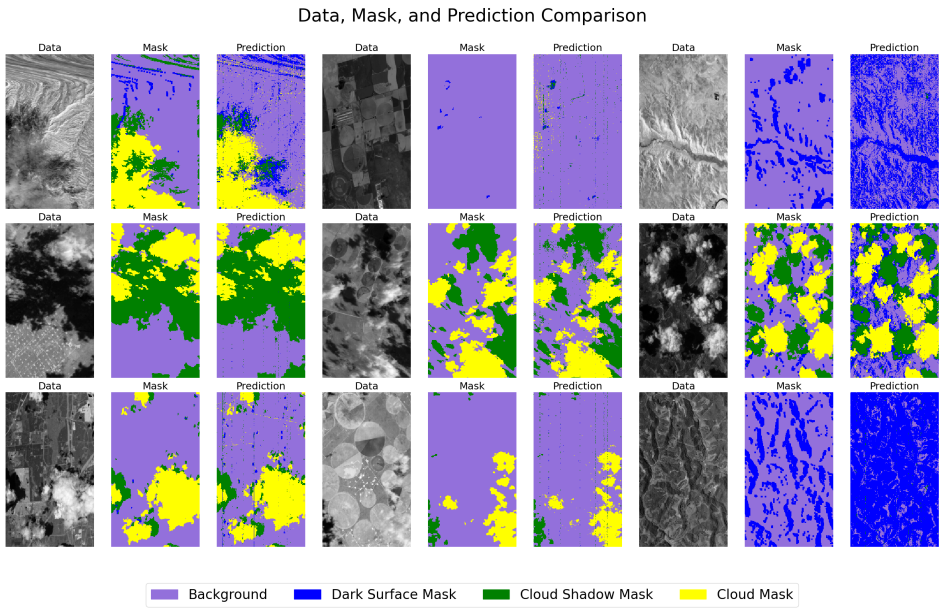


Figure 13: Radiance at 1592nm, ground truth labels, and predictions of the SCAN model for MethaneAIR data. The model shows enhanced boundary detection capabilities compared to U-Net, but with some residual noise in spectrally complex regions.

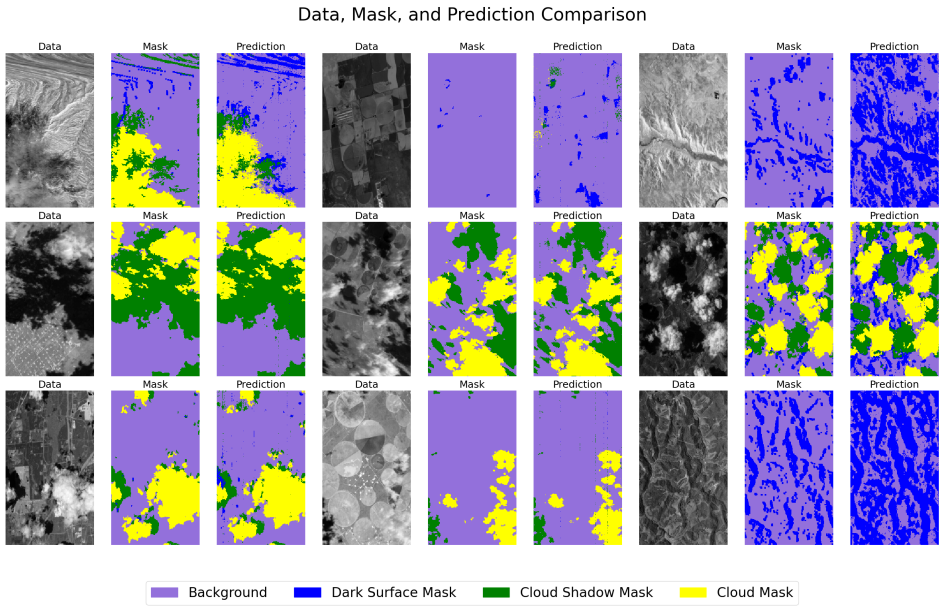


Figure 14: Radiance at 1592nm, ground truth labels, and predictions of the Combined MLP model for MethaneAIR data. The fusion approach demonstrates improved performance over individual models.

B.2 MethaneSAT Results

Figures 15 through 19 present detailed prediction results for all evaluated models on the MethaneSAT dataset. These visualizations complement the confusion matrices presented in Section 5 and demonstrate the performance of each model across diverse terrain and atmospheric conditions.

U-Net results (Figure 17) reveal significantly enhanced spatial coherence with the elimination of most noise artifacts. However, as noted in the main results, the model tends to undershoot the extent

of cloud formations in certain scenes and produces overly simplified shadow boundaries, limiting its overall accuracy.

The SCAN predictions (Figure 18) demonstrate notably improved performance compared to the U-Net model for MethaneSAT data, particularly in capturing the boundaries of cloud and shadow regions. This visual observation corresponds to the quantitative results presented in the main section, where SCAN outperformed U-Net on MethaneSAT data despite showing slightly lower performance on MethaneAIR.

The Combined MLP approach (Figure 19) shows enhanced performance through fusion, particularly in challenging scenes with varying surface reflectance. The model effectively reduces misclassification between spectrally similar classes while preserving the detailed structures of cloud formations.

Finally, the Combined CNN results (Figure 8) achieve the most accurate and balanced predictions across all MethaneSAT scenes. The model's ability to preserve spatial context during fusion enables it to effectively handle complex patterns of cloud and shadow formations while minimizing noise and misclassification, validating its superiority as demonstrated in the confusion matrices presented in the main results section.

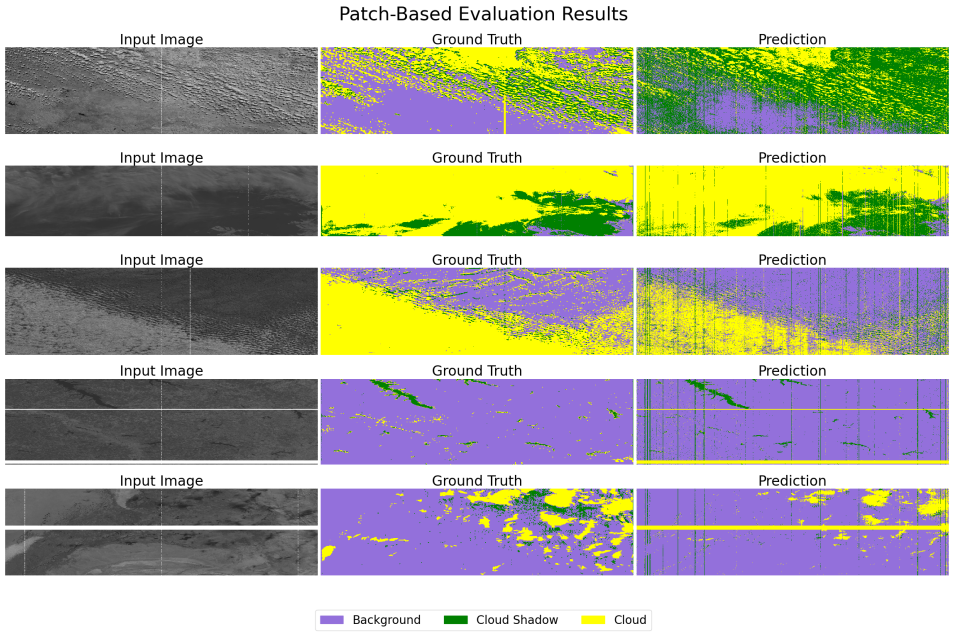


Figure 15: Data, ground truth labels, and predictions of the ILR model for MethaneSAT data. The predictions exhibit significant artifacts and striping patterns characteristic of sensor-specific noise.

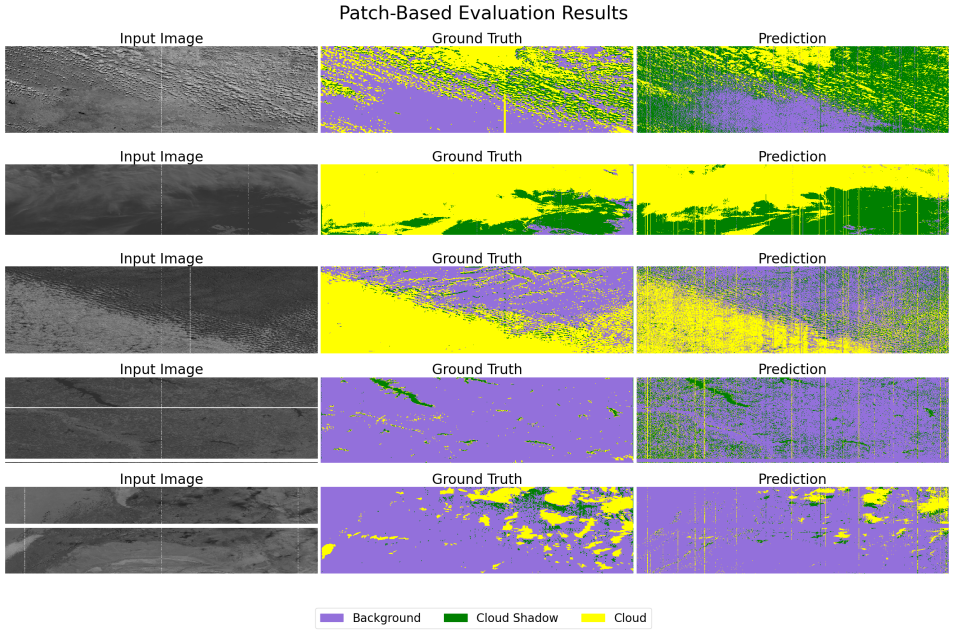


Figure 16: Data, ground truth labels, and predictions of the MLP model for MethaneSAT data. The model shows improved performance over ILR but retains horizontal striping artifacts.

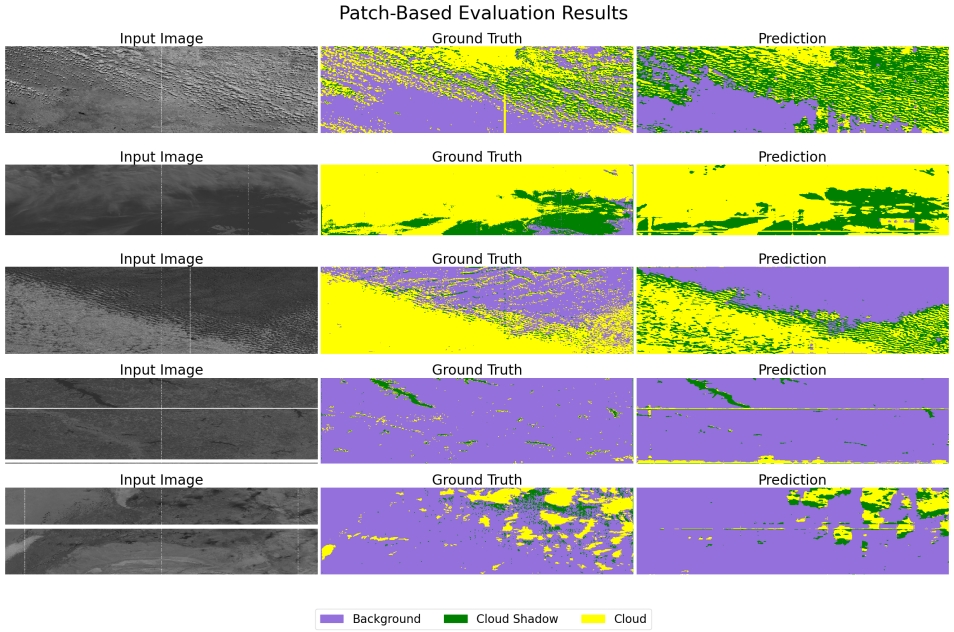


Figure 17: Data, ground truth labels, and predictions of the U-Net model for MethaneSAT data. The model produces spatially coherent predictions but tends to undershoot the extent of cloud formations in certain scenes.

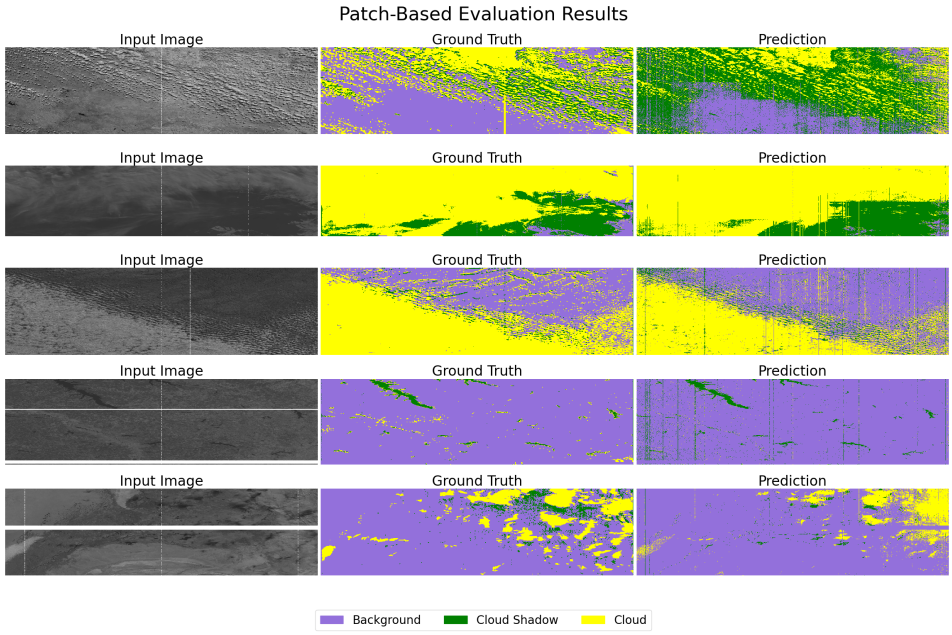


Figure 18: Data, ground truth labels, and predictions of the SCAN model for MethaneSAT data. The model demonstrates improved performance compared to U-Net, particularly in capturing the boundaries of cloud and shadow regions.

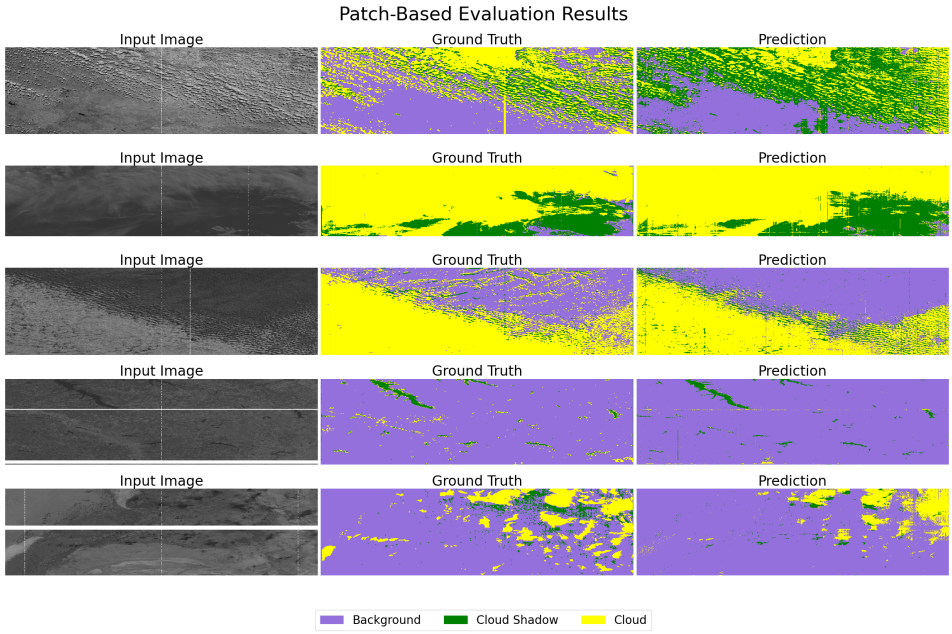


Figure 19: Data, ground truth labels, and predictions of the Combined MLP model for MethaneSAT data. The model shows enhanced performance through fusion, particularly in challenging scenes with varying surface reflectance.