University of Tehran

Earth Observation and Geomatics Engineering

Online ISSN: 2588-4360

Homepage: https://eoge.ut.ac.ir/

A Novel Hybrid Deep Learning Model for Methane Plume Detection in Oil and Gas Sectors Using PRISMA Satellite

Emadoddin Hemmati[1], and **Mehdi Mokhtarzade**[2]✉

1. Faculty of Geodesy and Geomatics Engineering & Remote Sensing Institute, K. N. Toosi University of Technology, Tehran, Iran. E-mail: aradhemmati@email.kntu.ac.ir

2. Corresponding author, Faculty of Geodesy and Geomatics Engineering & Remote Sensing Institute, K. N. Toosi University of Technology, Tehran, Iran. E-mail: m_mokhtarzade@kntu.ac.ir

Article Info

Article type:
Research Article

Article history:
Received 2026-01-26
Received in revised form 2026-06-27
Accepted 2026-06-30
Available online 2026-07-24

Keywords:
Methane,
Hyperspectral,
PRISMA,
Swin,
CNN

ABSTRACT

Accurate monitoring of methane (CH_4) emissions from oil and gas infrastructure is critical for near-term climate change mitigation. While hyperspectral satellites like PRISMA provide high-fidelity data, the optimal deep learning architecture for segmenting faint methane plumes remains undetermined. This study provides the first systematic comparison of convolutional and transformer-based models for this task.

We evaluate four architectures: an attention-enhanced U-Net++, Segformer, and two novel hybrids – SwinV2-UPP (SwinV2 encoder + U-Net++ decoder) and SwinV2-Former (SwinV2 encoder + MLP decoder). Models are trained on 73 manually delineated methane plumes from 34 PRISMA scenes across global oil and gas basins, augmented to 1,446 patches.

SwinV2-UPP achieved the highest F1-score of 0.834, a 4.0% improvement over the convolutional baseline (0.802). Notably, SwinV2-Former achieved state-of-the-art precision of 0.910 – outperforming SwinV2-UPP by 2.5% – demonstrating exceptional false-positive minimization. In zero-shot generalization, SwinV2-UPP maintained the highest F1 (0.725), while SwinV2-Former achieved the best precision (0.857).

We reveal a critical architectural trade-off: SwinV2-UPP balances precision and recall (F1=0.834) for comprehensive plume mapping, whereas SwinV2-Former prioritizes high-confidence detection (precision=0.910) for regulatory applications where false positives are costly. These findings provide a framework for selecting optimal deep learning architectures based on specific methane monitoring objectives.

Cite this article: Hemmati, E., Mokhtarzade, M. (2026). A Novel Hybrid Deep Learning Model for Methane Plume Detection in Oil and Gas Sectors Using PRISMA Satellite, *Earth Observation and Geomatics Engineering*, Volume 10, Issue 1, Pages 51-60. http//doi.org/ 10.22059/eoge.2026.410068.1214

Publisher: University of Tehran.

DOI: http//doi.org/ 10.22059/eoge.2026.410068.1214

1. Introduction

Industrialization ignited a surge in anthropogenic greenhouse gas emissions, primarily methane (CH4) and carbon dioxide (CO2), which have driven global warming through enhanced radiative forcing (Jones et al., 2023). This potent CH4 is expelled from myriad human activities spanning livestock operations, oil and gas infrastructure, coal extraction, waste management facilities, sewage treatment plants, and even rice agriculture (Collins et al., 2022). As CH4's second-largest contributor after CO2, accounting for 25% of warming since pre-industrial times, its relatively brief atmospheric residence time of 9-10 years renders mitigating its emissions a high-leverage strategy for near-term climate change mitigation (Karacan et al., 2024; Sobanaa et al., 2023a). A comprehensive understanding of these diverse emissions sources, coupled with rigorous monitoring and abatement measures targeting both long-lived CO2 and short-lived CH4, is imperative to restoring the planet's disrupted energy equilibrium and halting temperature ascension (Cael & Goodwin, 2023; Nunes, 2023; Sobanaa et al., 2023b).

Identifying and quantifying methane emissions, a pivotal greenhouse gas, from sources like landfills could pave the way for atmospheric concentration control and mitigation strategies (D. Cusworth et al., 2024; Da Silva et al., 2025). Cutting-edge developments in remote sensing capabilities have revolutionized our ability to observe and track methane releases globally from space-based platforms. Methane's unique spectral signatures necessitate specialized gas-sensitive detectors for identification. However, observations from satellites equipped with shortwave infrared (SWIR) bands offer potential for point source quantification (Huang et al., 2025; Kim et al., 2024; Meshcherinov et al., 2024; Zhang et al., 2023). The shortwave infrared region between around 1600-2500nm contains two absorption bands specific to methane, allowing measurements of solar radiation reflected from the Earth's surface in this part of the electromagnetic spectrum to detect methane (Jacob et al., 2022a; Tahmasbian et al., 2025; Varon et al., 2021a). Fortuitously, multispectral satellites like Sentinel-2 and Landsat-8, though unintended for methane monitoring, serendipitously capture methane's absorption bands. This coincidental sensitivity enables frequent mapping and columnar measurements of localized methane plumes despite suboptimal sensor configurations (Pandey et al., 2022; Varon et al., 2021b; Zhao et al., 2025). Nonetheless, hyperspectral instruments such as PRISMA and EnMAP yield more precise methane retrievals by their heightened ability to discern minute concentration variations. Yet, these advanced sensors are constrained by limited swath widths of just 30km and necessitate targeted acquisition requests, resulting in restricted data availability over broad areas compared to wider-ranging multispectral platforms (Guanter et al., 2021; Jacob et al., 2022b; Roger et al., 2024; Valverde et al., 2024). Guanter et al. (2021) focused on the retrieval of atmospheric methane (CH4) using PRISMA satellite observations (Guanter et al., 2021). They developed an algorithm specifically for PRISMA data and conducted a validation study comparing the retrieved CH4 values with other datasets. The study demonstrates the algorithm's effectiveness in accurately estimating CH4 concentrations and highlights the potential of PRISMA satellite observations for monitoring methane levels in the atmosphere. In addition, Cusworth et al. (2021) used PRISMA methane retrievals and Planet Labs satellite imagery to estimate methane emissions from Louisiana Haynesville shale gas production. The study aimed to provide accurate measurements of emissions from oil and gas operations in the region (G. Cusworth et al., 2021). Moreover, Varon et al. (2022) utilized high-resolution satellite observations and meteorological data to quantify methane point source emissions (Varon et al., 2022). By combining PRISMA methane retrievals with wind information, they detected localized methane plumes and calculated emission rates for the corresponding point sources. Jongaramrungruang et al. (2022) proposed MethaNet, a convolutional neural network (CNN) designed to quantify methane point-source emissions directly from high-resolution 2-D plume imagery without requiring ancillary wind speed measurements. Trained on large eddy simulation (LES) data with realistic noise over agricultural, desert, and urban environments, MethaNet achieved a mean absolute percentage error (MAPE) of 17% for emission rate prediction – a significant improvement over physics-based methods that rely on wind information. Validation against a controlled-release experiment confirmed agreement within precision error estimates. While MethaNet targets airborne AVIRIS-NG data (1–5 m resolution) rather than satellite observations, its core concept – using deep learning to infer emission rates directly from plume morphology – is directly relevant to our work. However, MethaNet addresses regression (emission rate quantification), whereas our study focuses on segmentation (plume detection and delineation), representing a complementary but distinct task.

Over the past few years, there has been remarkable progress in the field of deep learning, particularly with the rise of convolutional neural networks (CNNs). CNNs have proven to be highly successful models for image analysis, as they possess the ability to extract diverse features at different levels of abstraction (Chen et al., 2021; Sun et al., 2017; Tian, 2020). These networks are adept at capturing both fine-grained details and complex patterns, allowing for a comprehensive understanding of images. The versatility and effectiveness of CNNs have propelled their widespread adoption across various domains and applications. In this context, Joyce et al. (2022) utilized a CNN model trained on synthetic plume images to identify and quantify methane point source emissions from PRISMA hyperspectral satellite data (Joyce & Boesch, 2022). Their deep learning approach leveraged PRISMA's high resolution to pinpoint and estimate emission rates of localized methane hotspots from point sources like oil/gas operations and landfills. Vaughan et al. (2024) introduce CH4Net, a deep-learning model for monitoring methane super-emitters using Sentinel-2 data

(Vaughan et al., 2024). The model achieves high accuracy (0.83 scene-level, 0.77 pixel-level) and demonstrates effectiveness in identifying new super-emitter locations, with accuracies of 0.92 and 0.96, while also providing an open-source dataset of 925 methane plume masks. Most recently, Liu et al. (2025) proposed kMetha-Mamba, a novel architecture based on Mamba's selective state-space modeling, achieving 88.69% accuracy on AVIRIS-NG data with exceptionally low false positive rates (0.0466%), suggesting that state-space models may offer a promising alternative to both CNNs and transformers for methane plume segmentation.

While these studies have successfully demonstrated the efficacy of Convolutional Neural Networks (CNNs), the high dimensionality and complexity of hyperspectral data suggest that more advanced architectures may offer superior performance. This paper extends the existing body of work by conducting a systematic comparative analysis of state-of-the-art deep learning frameworks for methane plume segmentation in PRISMA imagery over oil and gas facilities. We investigate not only a robust convolutional baseline (Attention-Enhanced U-Net++) but also cutting-edge Vision Transformer (ViT) architectures (Segformer, SwinV2). The primary contributions of this work are twofold: first, the introduction of two novel hybrid architectures, SwinV2-UPP and SwinV2-Former, which synergize transformer encoders with advanced convolutional and MLP decoders; and second, a detailed analysis of the performance trade-offs between these different architectural paradigms. Our central hypothesis is that the global self-attention mechanisms inherent to transformers can more effectively model the diffuse, complex spatial-spectral patterns of methane plumes than locally-focused CNNs. By comparing these distinct frameworks, we aim to identify the optimal architecture for precise, automated methane detection from high-fidelity hyperspectral data.

2. Material and methods

Figure 1. An overview of the research workflow employed in this study. The methodology is organized into three principal stages: (1) data preparation, which involves identifying potential emission sites from public sources, processing satellite imagery to retrieve methane concentrations, and generating an augmented training dataset through manual delineation; (2) Model Development, where four distinct deep learning architectures are constructed by systematically pairing state-of-the-art transformer and convolutional components; and finally, (3) Training and Assessment, in which the models are trained and their segmentation performance is quantitatively and qualitatively evaluated.

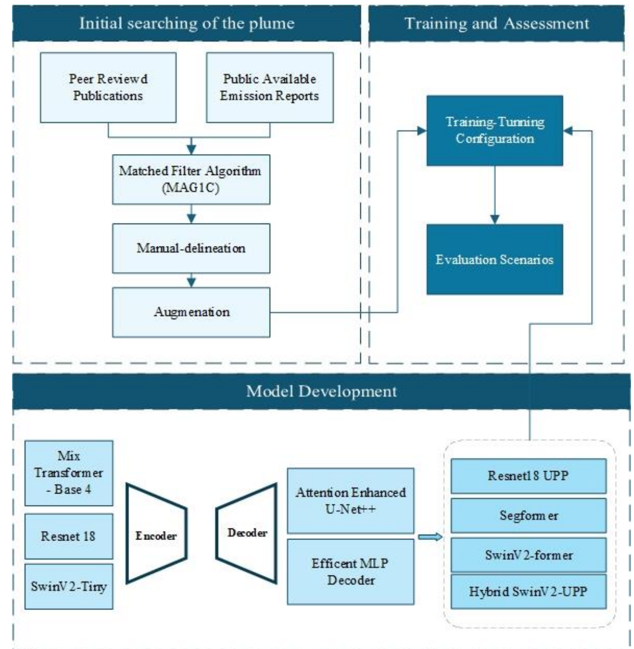

Figure 1. shows the overall flowchart of this study

2.1. Dataset

This investigation leverages hyperspectral data from the PRecursore IperSpettrale della Missione Applicativa (PRISMA) satellite to analyse methane (CH4) plumes from oil and gas production facilities. Launched by the Italian Space Agency in March 2019, PRISMA provides publicly available, targeted imagery at a 30 m × 30 m spatial resolution, with each scene covering a 30 km × 30 km area. Its instrument captures a broad spectral range from 400 to 2500 nm with a resolution of approximately 10 nm, making it exceptionally well-suited for point-source methane detection (Irakulis-Loitxate et al., 2022a; Joyce et al., 2023).

The satellite's utility for this task is primarily due to its sensor's high signal-to-noise ratio (SNR) and its sensitivity across methane's key shortwave infrared (SWIR) absorption bands—a weaker feature at 1650-1750 nm and a stronger one at 2200-2400 nm, as illustrated in Figure. 1. Our study targets major hydrocarbon-producing basins known for significant methane emissions, utilizing top-of-atmosphere (TOA) radiance datasets from PRISMA. The geographical distribution of the PRISMA scenes acquired is depicted in Figure 2, highlighting key infrastructure in the Permian Basin (United States), Korpeje (Turkmenistan), and various sites across the Middle East. Initial site identification was guided by peer-reviewed literature and publicly available emission reports from sources such as the GHGSat Spectra website[1] (Irakulis-Loitxate et al., 2022b; Schuit et al., 2023).

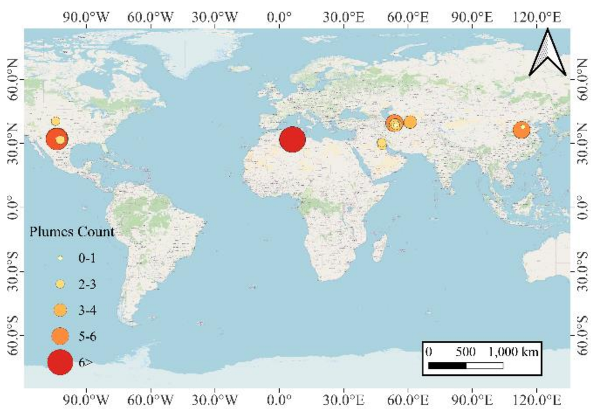

Figure 2. show distribution detected methane in Oil and Gas sectors around the world

2.1. Methane Retrieval and Training Data Generation

A high-quality, labeled dataset is foundational for training a supervised deep learning model. The process of creating our training data began by converting the raw PRISMA radiance data into methane concentration maps. For this, we employed the robust matched filter (MF) algorithm, a technique widely applied in remote sensing for target detection (Thorpe et al., 2013; Thompson et al., 2015; Irakulis-Loitxate et al., 2021). Our pipeline specifically incorporates the MAG1C (Matched filter with Albedo correction and reweighted L1 sparsity Code) implementation, whose performance has been previously validated for PRISMA data (Irakulis-Loitxate et al., 2021). The retrieval is governed by Equation (1):

$$\alpha = \frac{(x-\mu)^T \Sigma^{-1} t}{t^T \Sigma^{-1} t} \tag{1}$$

In this equation, α represents the retrieved CH4 concentration in parts-per-billion (ppb), is the spectrum of the pixel being analyzed, is the target CH4 spectrum, while and denote the mean and covariance of the background radiance, respectively (Guanter et al., 2021b; Roger, Guanter, et al., 2024).

Each resulting concentration map was then visually inspected to identify distinct methane plumes. Plumes that met our criteria were manually delineated using polygons, with boundaries drawn in consideration of prevailing wind directions. To ensure the model learns spectral features rather than spurious surface correlations, we applied strict exclusion criteria, discarding any plumes that were clearly visible in standard RGB imagery or coincided with obvious albedo artifacts. The confirmed plume regions were then extracted as 112 × 112 patches.

34 images were confirmed to contain distinct methane plumes suitable for model training. To mitigate the data scarcity inherent in targeted satellite acquisitions and to enhance model generalization, we implemented two data augmentation strategies. The first technique applied random rotations (90°, 180°, 270°) to the image patches. The second involved shifting the window to create a patch that plume is not only in the center. These augmentation procedures expanded our initial set of 73 unique plumes into a final training dataset of 1446 patches.

2.3. Methodology

The semantic segmentation of methane plumes from hyperspectral imagery presents a unique challenge, requiring models that can interpret high-dimensional data and capture faint, large-scale spatial patterns. Our methodology is centred on a comparative analysis of deep learning models, systematically adapting and hybridizing state-of-the-art convolutional and transformer architectures to this specific task.

2.3.1. Foundational Architectures

The dataset is divided into training (70%) and testing (30%) Our investigation leverages several powerful, publicly recognized architectures as foundational components, which have been customized to our specific task.

- Swin Transformer (SwinV2): This vision transformer architecture computes self-attention within non-overlapping, shifted local windows. This design achieves high efficiency and focuses on hierarchical feature extraction, capturing both fine-grained local details and global contextual information as the windows are shifted across layers. Its capacity for modelling long-range dependencies is particularly hypothesized to be effective for detecting the diffuse nature of methane plumes (Liu et al., 2021).
- Mix Transformer (MiT-B4) and Segformer: The Segformer model consists of a Mix Transformer (MiT) encoder and a lightweight MLP decoder. The MiT encoder is a hierarchical transformer that generates multi-scale features without requiring positional encoding, making it computationally efficient. The simple MLP decoder then aggregates these features, providing a powerful yet fast segmentation framework (Xie et al., 2021).
- U-Net++: This is an advanced convolutional architecture characterized by nested, dense skip connections. These pathways aim to reduce the semantic gap between encoder and decoder feature maps, enabling the aggregation of features at multiple scales and leading to more precise segmentation, especially at object boundaries (Mohammadi, 2025).

2.3.2. Model configurations and novelty

Our primary novelty in model architecture lies in the strategic hybridization of these components and the integration of attention mechanisms to create models tailored for methane detection. We implemented and evaluated four distinct configurations:

1. ResNet18 UPP: To establish a strong convolutional baseline, we employed a U-Net++ model with a ResNet-18 encoder. We augmented its decoder with a Spatial and Channel Squeeze & Excitation (scSE) attention module. This addition allows the network to dynamically recalibrate feature maps, emphasizing salient spectral and spatial features relevant to methane while suppressing background noise.
2. Standard Segformer (MiT-B4): We utilized the Segformer architecture with a MiT-B4 encoder as a robust transformer-based baseline. Its inherent efficiency and multi-scale feature fusion provide a strong benchmark for transformer performance on our task.
3. Hybrid SwinV2-UPP: Our first novel hybrid model combines the powerful SwinV2 encoder with the dense, feature-aggregating U-Net++ decoder (also enhanced with scSE attention). This configuration tests the hypothesis that a state-of-the-art transformer encoder can enhance the performance of a proven convolutional decoder by providing it with a richer, more context-aware set of initial features.
4. Proposed SwinV2-Former Architecture: Our principal proposed architecture, the SwinV2-Former, couples the SwinV2 encoder with the lightweight Segformer MLP decoder. We posit this is the most effective combination. It leverages the superior hierarchical feature representation and global context modelling of the SwinV2 transformer while benefiting from the computational efficiency and effective feature fusion of the Segformer decoder. This synergy is expected to yield a model that is not only highly accurate but also computationally tractable for large-scale remote sensing applications.

2.3.3. Training protocol and loss function

2.3.3.1. Optimization Strategy

We employed the Adam optimizer, a widely adopted and effective algorithm for deep learning applications, with an initial learning rate of 0.0001. To facilitate stable convergence and fine-tuning, the learning rate was dynamically managed by a ReduceLROnPlateau scheduler. This scheduler monitors the F1-score on the validation set and reduces the learning rate by a factor of 0.5 if no improvement is observed after 5 consecutive epochs. This strategy allows the model to make large initial progress and then take smaller, more precise steps as it approaches an optimal solution.

2.3.3.2. Composite loss function

The segmentation of methane plumes is challenged by significant class imbalance (plumes constitute a small fraction of pixels) and often indistinct boundaries. To address this, we designed a composite loss function that combines three distinct loss functions, each targeting a different aspect of the segmentation task. The total loss () is a weighted sum of Dice Loss, Focal Loss, and standard Binary Cross-Entropy.

The final loss is calculated a s:

$$L_{Total} = 0.3L_{Dice} + 0.4L_{Focal} + 0.3L_{BCE} \tag{2}$$

Dice Loss: This loss is an overlap-based metric that directly optimizes the F1-score, making it highly effective for imbalanced datasets. It evaluates the spatial overlap between the predicted mask and the ground truth. It is defined as:

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i + \varepsilon}{\sum_{i=1}^{N} p_i + \sum_{1}^{N} g_i + \varepsilon} \tag{3}$$

where is the predicted probability for pixel, is the ground truth label (0 or 1), is the total number of pixels, and is a small smoothing constant to prevent division by zero.

Focal Loss: Originally proposed for object detection, Focal Loss is an enhancement of the standard cross-entropy loss that addresses class imbalance by down-weighting the loss assigned to well-classified examples. This forces the model to focus its learning on harder, misclassified examples, such as faint plume edges. It is formulated as:

$$L_{Focal} = -\frac{1}{N}\sum_{i=1}^{N} [\alpha\left(1 - p_i\right)^{\gamma} g_i \log(p_i) + (1-\alpha)p_i^{\gamma}(1 - g_i)] \tag{4}$$

where γ is the focusing parameter (we use γ = 2) that adjusts the rate at which easy examples are down-weighted, and α is a balancing parameter.

Binary Cross-Entropy Loss: This is the standard pixel-wise logarithmic loss for binary classification tasks. It provides a stable and reliable gradient for the fundamental task of classifying each pixel as either methane plume or background. The formula is:

$$L_{BCE} = \frac{1}{N}\sum_{i=1}^{N} [(g_i \log p_i) + (1 - p_i)(1 - g_i)] \tag{5}$$

By combining these three functions with a higher weight on Focal Loss, our training objective robustly handles the class imbalance, improves the localization of plume boundaries, and ensures stable convergence.

3. Result

The quantitative performance of the four evaluated model architectures across different spectral band configurations is presented in Table 1. The models were evaluated using F1-score, Precision, and Recall on both a standard test set and a zero-shot generalization set.

Overall, the models incorporating the SwinV2 encoder demonstrated superior performance. The Hybrid SwinV2-UPP model achieved the highest overall F1-score of 0.834, registered when utilizing only the SWIR2 band subset. This was closely followed by the SwinV2-Former architecture, which obtained a peak F1-score of 0.828, also with the SWIR2 bands.

A key finding emerged in the precision metric, where the SwinV2-Former model significantly outperformed all other architectures, achieving a state-of-the-art precision of 0.910. The Segformer model served as a strong transformer-based baseline, attaining a commendable F1-score of 0.827 with the SWIR2 bands. The attention-enhanced U-Net++ model, representing the pure convolutional approach, recorded a maximum F1-score of 0.802 when using all available hyperspectral bands.

The choice of spectral bands had a notable impact on model performance. For the three highest-performing models (Hybrid SwinV2-UPP, SwinV2-Former, and Segformer), the most effective configurations utilized focused band subsets (primarily SWIR2), rather than the complete hyperspectral cube. In contrast, the baseline U-Net++ model achieved its best result with all 239 bands.

In the zero-shot evaluation, all models exhibited a predictable decrease in performance, highlighting the challenge of domain generalization. The Hybrid SwinV2-UPP model maintained the highest F1-score in this scenario (0.725), though the SwinV2-Former demonstrated robust generalization when supplied with both SWIR-1 and SWIR-2 bands (0.717).

Table 1. Comparative Performance of Model Architectures and Band Selections

	Test			**Zero Shot**		
	SWIR 2	**All Bands**	**SWIR-1 SWIR-2**	**SWIR 2**	**All Bands**	**SWIR-1 SWIR-2**
ResNet18 UPP						
F1	0.774	0.802	0.783	0.658	0.682	0.710
Precision	0.840	0.829	0.816	0.711	0.765	0.830
Recall	0.717	0.777	0.752	0.612	0.616	0.620
Segformer						
F1	0.827	0.791	**0.811**	0.701	0.683	0.715
Precision	0.882	0.874	0.867	0.777	0.786	0.814
Recall	0.778	0.723	**0.761**	**0.638**	0.604	**0.637**
Hybrid SwinV2-UPP						
F1	**0.834**	**0.824**	0.799	0.697	**0.720**	0.714
Precision	0.885	0.861	0.859	0.805	0.829	0.821
Recall	**0.789**	**0.790**	0.746	0.614	**0.636**	0.632
SwinV2-Former						
F1	0.828	0.803	0.810	0.688	0.698	**0.717**
Precision	**0.910**	**0.904**	**0.894**	**0.832**	**0.846**	**0.857**
Recall	0.760	0.723	0.741	0.587	0.594	0.616

Figure 3, highlights significant variability in their ability to accurately identify multiple regions of interest within a single scene. The Hybrid SwinV2-UPP model (C) demonstrates the highest fidelity to the ground truth, successfully segmenting both the larger, elongated structure and the smaller, crescent-shaped object with clean boundaries. In contrast, the other models exhibit partial detection failure. While Segformer (A) accurately delineates the larger structure, it completely fails to identify multiple regions of interest within a single scene. The Hybrid SwinV2-UPP model (C) demonstrates the highest fidelity to the ground truth, successfully segmenting both the larger, elongated structure and the smaller, crescent-shaped object with clean boundaries. In contrast, the other models exhibit partial detection failure. While Segformer (A) accurately delineates the larger structure, it completely fails to identify the smaller one. Conversely, SwinV2-Former (B) successfully segments the smaller object but overlooks the larger structure entirely. The ResNet18 UPP model (D) produces the most divergent result, characterized by a highly fragmented and noisy prediction for the primary object and a complete failure to detect the secondary object. This visual evidence underscores the superior capability of the Hybrid SwinV2-UPP architecture in handling scenes with multiple, disconnected objects of varying scales in this specific instance.

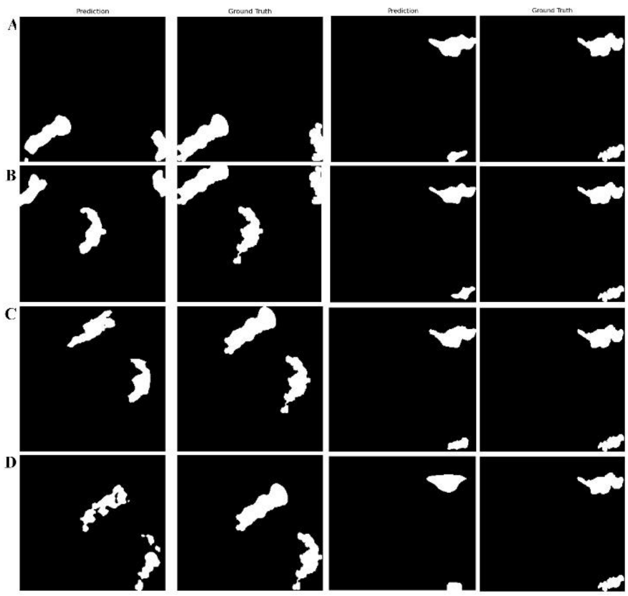

Figure 3. offers a qualitative comparison of segmentation performance between four deep learning models: Segformer (A), SwinV2-Former (B), Hybrid SwinV2-UPP (C), and ResNet18 UPP (D).

4. Discussion

The experimental results provide several key insights into the architectural requirements for effective methane plume segmentation from hyperspectral data. Our findings confirm the central hypothesis that advanced transformer-based models, particularly those leveraging the SwinV2 encoder, offer a significant performance advantage over purely convolutional approaches. However, the results also reveal a nuanced and unexpected trade-off between architectural design and performance metrics.

4.1. Architectural determinants of segmentation precision: contextual awareness vs. boundary fidelity

A key and somewhat counter-intuitive finding were the superior precision of the SwinV2-Former architecture over the Hybrid SwinV2-UPP. While U-Net architectures are renowned for high-precision boundary delineation, our results suggest that for diffuse, low-contrast targets like methane plumes, the primary challenge for precision shifts. Rather than meticulously refining a well-defined edge, the task becomes the confident identification of the faint signal itself from a noisy background.

Herein lies the architectural divergence. The convolutional U-Net++ decoder operates on a more local receptive field. It excels at boundary refinement but can be more easily misled by localized spectral artifacts, potentially leading to false positives. In contrast, the Swin Transformer's self-attention mechanism is explicitly designed to model long-range dependencies and build a global, contextual understanding of the entire scene. This ability to build contextual confidence appears to be more effective at filtering out false positives for this specific problem, resulting in the SwinV2-Former 's state-of-the-art precision.

4.2. Synergistic performance of the hybrid transformer-CNN architecture

Despite the SwinV2-Former's superior precision, the hybrid SwinV2-UPP model delivered the highest overall F1-score. This indicates a better balance between precision and recall, suggesting it successfully identified more true plumes. The strength of this hybrid model aligns with a growing body of work demonstrating that combining transformer encoders with CNN-based decoders can leverage the strengths of both paradigms.

The SwinV2 encoder provides the global context necessary to identify large, diffuse plumes. The U-Net++ decoder, with its dense skip connections, then leverages these rich feature maps to excel at its inherent strength: precisely delineating boundaries. The hybrid SwinV2-UPP, therefore, succeeds as the best "all-rounder" by combining a transformer's global view with a convolutional network's local specialization, making it robust to variations in plume size and concentration.

4.3. The impact of spectral dimensionality on decoder performance

A significant finding was the detrimental effect of using all available bands on the MLP-based decoders (Segformer and SwinV2-Former). Both models saw a notable drop in performance when moving from a focused SWIR subset to the full 239-band hyperspectral cube. This phenomenon, where an excess of features can degrade the performance of certain classifiers, is a well-documented challenge in hyperspectral analysis, often referred to as the Hughes phenomenon **[35]**. Our results suggest that the convolutional filters and feature aggregation pathways in U-Net++ are inherently more robust to this high dimensionality. This finding has practical implications, suggesting that for efficient architectures with lightweight decoders, a deliberate band selection strategy is not just a computational shortcut but a requirement for optimal performance.

4.4. Model generalization and future research directions

The performance drop in the zero-shot scenario was expected and underscores the challenge of domain shift in remote sensing applications. However, the observation that the SwinV2-Former generalized best when using both SWIR-1 and SWIR-2 bands is noteworthy. It suggests that while a SWIR-1 may be optimal for a specific data distribution, a richer feature set including both absorption windows may be key to building more robust and generalizable models for global monitoring.

In conclusion, our work demonstrates that hybrid transformer-CNN architectures, particularly the Hybrid SwinV2-UPP, provide a state-of-the-art balance of precision and recall for methane plume segmentation. The SwinV2-Former, however, emerges as a specialist model for high-confidence applications where minimizing false positives is the primary objective. Future work should explore dynamic, attention-based band selection mechanisms and investigate the training of specialist models for plumes of varying sizes to potentially combine their strengths in an operational ensemble.

5. Conclusion

This study evaluated deep learning models for methane plume segmentation from PRISMA data, demonstrating that hybrid architectures using the SwinV2 transformer encoder significantly outperform pure convolutional approaches. Our novel SwinV2-UPP and SwinV2-Former models yielded state-of-the-art results, revealing a critical trade-off between architectural design and monitoring objectives.

The SwinV2-UPP, pairing the transformer with a U-Net++ decoder, provided the best all-around performance with the highest F1-score, making it ideal for comprehensive plume mapping. In contrast, the SwinV2-Former, using a lightweight MLP decoder, achieved state-of-the-art precision, establishing it as the superior choice for high-confidence detection where minimizing false positives is the priority. This work also underscores the importance of strategic band selection for optimizing performance. Ultimately, these findings provide a clear framework for operators to select the optimal deep learning tool based on their specific operational needs, whether for balanced detection or high-precision verification.

References

Cael, B., & Goodwin, P. (2023). Global methane pledge versus carbon dioxide emission reduction. *Environmental Research Letters*, *18*. (https://doi.org/10.1088/1748-9326/acf8dd)

Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., & Miao, Y. (2021). Review of Image Classification Algorithms Based on Convolutional Neural Networks. *Remote. Sens.*, *13*, 4712. (https://doi.org/10.3390/rs13224712)

Collins, W., Orbach, R., Bailey, M., Biraud, S., Coddington, I., DiCarlo, D., Peischl, J., Radhakrishnan, A., & Schimel, D. (2022). Monitoring methane emissions from oil and gas operations‡. *Optics Express*, *30 14*, 24326–24351. (https://doi.org/10.1364/oe.464421)

Cusworth, D., Duren, R., Ayasse, A., Jiorle, R., Howell, K., Aubrey, A., Green, R., Eastwood, M., Chapman, J., Thorpe, A., Heckler, J., Asner, G., Smith, M., Thoma, E., Krause, M., Heins, D., & Thorneloe, S. (2024). Quantifying methane emissions from United States landfills. *Science*, *383*, 1499–1504. (https://doi.org/10.1126/science.adi7735)

Cusworth, G., Garnett, T., & Lorimer, J. (2021). Legume dreams: The contested futures of sustainable plant-based food systems in Europe. *Global Environmental Change*, *69*. (https://doi.org/10.1016/j.gloenvcha.2021.102321)

Da Silva, T. A., Andrade, I., Araujo, E., Correa, T., Mendes, A. C. P., Ribeiro, F. N. D., Andrade, M., & Landulfo, E. (2025). Landfill methane emission: a case study using inversion methods, satellite and CRDS-based observations. *Frontiers in Earth Science*. (https://doi.org/10.3389/feart.2025.1549385)

Guanter, L., Irakulis-Loitxate, I., Gorroño, J., Sánchez-García, E., Cusworth, D., Varon, D., Cogliati, S., & Colombo, R. (2021a). Mapping methane point emissions with the PRISMA spaceborne imaging spectrometer. *Remote Sensing of Environment*. (https://doi.org/10.1016/j.rse.2021.112671)

Guanter, L., Irakulis-Loitxate, I., Gorroño, J., Sánchez-García, E., Cusworth, D., Varon, D., Cogliati, S., & Colombo, R. (2021b). Mapping methane point emissions with the PRISMA spaceborne imaging spectrometer. *Remote Sensing of Environment*. (https://doi.org/10.1016/j.rse.2021.112671)

Huang, Y., Han, G., Shi, T., Li, S., Mao, H., Nie, Y., & Gong, W. (2025). FI-SCAPE: A Divergence Theorem Based Emission Quantification Model for Air/Spaceborne Imaging Spectrometer Derived XCH4 Observations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *18*, 255–272. (https://doi.org/10.1109/JSTARS.2024.3490896)

Irakulis-Loitxate, Itziar & Guanter, Luis & Maasakkers, Joannes & Zavala-Araiza, Daniel & Aben, I.. (2022). Satellites Detect Abatable Super-Emissions in One of the World's Largest Methane Hotspot Regions. Environmental science & technology. 56. (https://doi.org/10.1021/acs.est.1c04873.)

Jacob, D., Varon, D., Cusworth, D., Dennison, P., Frankenberg, C., Gautam, R., Guanter, L., Kelley, J., Mckeever, J., Ott, L., Poulter, B., Qu, Z., Thorpe, A., Worden, J., & Duren, R. (2022a). Quantifying methane emissions from the global scale down to point sources using satellite observations of atmospheric methane. *Atmospheric Chemistry and Physics*. (https://doi.org/10.5194/acp-22-9617-2022)

Jacob, D., Varon, D., Cusworth, D., Dennison, P., Frankenberg, C., Gautam, R., Guanter, L., Kelley, J., Mckeever, J., Ott, L., Poulter, B., Qu, Z., Thorpe, A., Worden, J., & Duren, R. (2022b). Quantifying methane emissions from the global scale down to point sources using satellite observations of atmospheric methane. *Atmospheric Chemistry and Physics*. (https://doi.org/10.5194/acp-22-9617-2022)

Jones, M., Peters, G., Gasser, T., Andrew, R., Schwingshackl, C., Gütschow, J., Houghton, R., Friedlingstein, P., Pongratz, J., & Quéré, L. (2023). National contributions to climate change due to historical emissions of carbon dioxide, methane, and nitrous oxide since 1850. *Scientific Data*, *10*. (https://doi.org/10.1038/s41597-023-02041-1)

Jongaramrungruang, S., Andrew K. Thorpe, Georgios Matheou, Christian Frankenberg, MethaNet – An AI-driven approach to quantifying methane point-source emission from high-resolution 2-D plume imagery, Remote Sensing of Environment, Volume 269, 2022, 112809, ISSN 0034-4257, (https://doi.org/10.1016/j.rse.2021.112809.)

Joyce, Peter & Villena, Cristina & Huang, Yahui & Webb, Alex & Gloor, Manuel & Wagner, Fabien H & Chipperfield, Martyn & Barrio Guilló, Rocío & Wilson, Chris & Boesch, Hartmut. (2023). Using a deep neural network to detect methane point sources and quantify emissions from PRISMA hyperspectral satellite images. Atmospheric Measurement Techniques. 16. 2627-2640. (https://doi.org/10.5194/amt-16-2627-2023.)

Karacan, C., Field, R., Olczak, M., Kasprzak, M., Ruiz, F., & Schwietzke, S. (2024). Mitigating climate change by abating coal mine methane: A critical review of status and opportunities. *International Journal of Coal Geology*. (https://doi.org/10.1016/j.coal.2024.104623)

Kim, M., Park, J., & Hyun, C.-U. (2024). Current Status of Satellite Remote Sensing-Based Methane Emission Monitoring Technologies. *Economic and Environmental Geology*. (https://doi.org/10.9719/eeg.2024.57.5.513)

Liu, Y., Hailiang Shi, Ke Cao, Shichao Wu, Hanhan Ye, Xianhua Wang, Erchang Sun, Yunfei Han, Wei Xiong, kMetha-Mamba: K-means clustering mamba for methane plumes segmentation, International Journal of Applied Earth Observation and Geoinformation, Volume 142, 2025, 104664, ISSN 1569-8432,(https://doi.org/10.1016/j.jag.2025.104664.)

Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 9992-10002, (https://doi.org/10.1109/ICCV48922.2021.00986.)

Meshcherinov, V., Kazakov, V., Spiridonov, M., Suvorov, G., & Rodin, A. (2024). Lidar-based gas analyzer for remote sensing of atmospheric methane. *Sensors and Actuators B: Chemical*. (https://doi.org/10.1016/j.snb.2024.136899)

Mohammadi, J. (2025). *A Comparative Study of U-Net, Attention U-Net and U-Net++ for Brain MRI Segmentation: Advancing Automated Tumor Detection to Streamline Clinical Diagnostics*.

Nunes, L. (2023). The Rising Threat of Atmospheric CO2: A Review on the Causes, Impacts, and Mitigation Strategies. *Environments*. (https://doi.org/10.3390/environments10040066)

Pandey, S., Van Nistelrooij, M., Maasakkers, J., Sutar, P., Houweling, S., Varon, D., Tol, P., Gains, D., Worden, J., & Aben, I. (2022). Daily detection and quantification of methane leaks using Sentinel-3: a tiered satellite observation approach with Sentinel-2 and Sentinel-5p. *Remote Sensing of Environment*. (https://doi.org/10.1016/j.rse.2023.113716)

Roger, J., Guanter, L., Gorroño, J., & Irakulis-Loitxate, I. (2024). Exploiting the entire near-infrared spectral range to improve the detection of methane plumes with high-resolution imaging spectrometers. *Atmospheric Measurement Techniques*. (https://doi.org/10.5194/amt-17-1333-2024)

Roger, J., Irakulis-Loitxate, I., Valverde, A., Gorroño, J., Chabrillat, S., Brell, M., & Guanter, L. (2024). High-Resolution Methane Mapping With the EnMAP Satellite Imaging Spectroscopy Mission. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–12. (https://doi.org/10.1109/TGRS.2024.3352403)

Schuit, B. J., Maasakkers, J. D., Bijl, P., Mahapatra, G., Van den Berg, A.-W., Pandey, S., Lorente, A., Borsdorff, T., Houweling, S., & Varon, D. J. (2023). Automated detection and monitoring of methane super-emitters using satellite data. *Atmospheric Chemistry and Physics Discussions*, *2023*, 1–47. (https://doi.org/10.5194/acp-23-9071-2023.)

Sobanaa, M., Prathiviraj, R., Selvin, J., & Prathaban, M. (2023a). A comprehensive review on methane's dual role: effects in climate change and potential as a carbon–neutral energy source. *Environmental Science and Pollution Research*, 1–16. (https://doi.org/10.1007/s11356-023-30601-w)

Sobanaa, M., Prathiviraj, R., Selvin, J., & Prathaban, M. (2023b). A comprehensive review on methane's dual role: effects in climate change and potential as a carbon–neutral energy source. *Environmental Science and Pollution Research*, 1–16. (https://doi.org/10.1007/s11356-023-30601-w)

Sun, Y., Xue, B., Zhang, M., & Yen, G. (2017). Evolving Deep Convolutional Neural Networks for Image Classification. *IEEE Transactions on Evolutionary Computation*, *24*, 394–407. (https://doi.org/10.1109/TEVC.2019.2916183)

Tahmasbian, I., Navas, A., & Dunlop, M. (2025). Feature Wavelengths for Quantifying Methane Concentrations Using Shortwave Infrared Hyperspectral Imaging: A Controlled Condition Study. *Analytical Chemistry*. (https://doi.org/10.1021/acs.analchem.4c05590)

Tian, Y. (2020). Artificial Intelligence Image Recognition Method Based on Convolutional Neural Network Algorithm. *IEEE Access*, *8*, 125731–125744. (https://doi.org/10.1109/ACCESS.2020.3006097)

Valverde, A., Roger, J., Gorroño, J., Irakulis-Loitxate, I., & Guanter, L. (2024). Detecting methane emissions from palm oil mills with airborne and spaceborne imaging spectrometers. *Environmental Research Letters*, *19*. (https://doi.org/10.1088/1748-9326/ad8806)

Varon, D., Jacob, D., Sulprizio, M., Estrada, L., Downs, W., Shen, L., Hancock, S., Nesser, H., Qu, Z., Penn, E., Chen, Z., Lu, X., Lorente, A., Tewari, A., & Randles, C. (2022). Integrated Methane Inversion (IMI 1.0): a user-friendly, cloud-based facility for inferring high-resolution methane emissions from TROPOMI satellite observations. *Geoscientific Model Development*. (https://doi.org/10.5194/gmd-15-5787-2022)

Varon, D., Jervis, D., Mckeever, J., Spence, I., Gains, D., & Jacob, D. (2021a). High-frequency monitoring of anomalous methane point sources with multispectral Sentinel-2 satellite observations. *Atmospheric Measurement Techniques*, *14*, 2771–2785. (https://doi.org/10.5194/AMT-14-2771-2021)

Varon, D., Jervis, D., Mckeever, J., Spence, I., Gains, D., & Jacob, D. (2021b). High-frequency monitoring of anomalous methane point sources with multispectral Sentinel-2 satellite observations. *Atmospheric Measurement Techniques*, *14*, 2771–2785. (https://doi.org/10.5194/AMT-14-2771-2021)

Vaughan, A., Mateo-García, G., Gómez-Chova, L., Ruzicka, V., Guanter, L., & Irakulis-Loitxate, I. (2024). CH4Net: a deep learning model for monitoring methane super-emitters with Sentinel-2 imagery. *Atmospheric Measurement Techniques*. (https://doi.org/10.5194/amt-17-2583-2024)

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, *34*, 12077–12090. (https://doi.org/10.48550/arXiv.2105.15203)

Zhang, S., Jun, Zhang, X., & Guo, C. (2023). Atmospheric remote sensing for anthropogenic methane emissions: Applications and research opportunities. *The Science of the Total Environment*, 164701. (https://doi.org/10.1016/j.scitotenv.2023.164701)

Zhao, S., Zhang, Y., Zhao, S., Wang, X., & Varon, D. (2025). A data-efficient deep transfer learning framework for methane super-emitter detection in oil and gas fields using the Sentinel-2 satellite. *Atmospheric Chemistry and Physics*. (https://doi.org/10.5194/acp-25-4035-2025)