Robust Small Methane Plume Segmentation in Satellite Imagery

Khai Duc Minh Tran[1], Hoa Van Nguyen[1], Aimuni Binti Muhammad Rawi[2], Hareeshrao Athinarayanarao[2], Ba-Ngu Vo[1]

Abstract—This paper tackles the challenging problem of detecting methane plumes, a potent greenhouse gas, using Sentinel-2 imagery. This contributes to the mitigation of rapid climate change. We propose a novel deep learning solution based on U-Net with a ResNet34 encoder, integrating dual spectral enhancement techniques (Varon ratio and Sanchez regression) to optimise input features for heightened sensitivity. A key achievement is the ability to detect small plumes down to 400 m² (i.e., for a single pixel at 20 m resolution), surpassing traditional methods limited to larger plumes. Experiments show our approach achieves a 78.39% F1-score on the validation set, demonstrating superior performance in sensitivity and precision over existing remote sensing techniques for automated methane monitoring, especially for small plumes.

I. INTRODUCTION

Methane (CH_4) is the second most influential anthropogenic greenhouse gas after carbon dioxide, responsible for approximately a quarter of observed global warming since pre-industrial times [1–3]. Due to its relatively short atmospheric lifetime of about 9-12 years, achieving rapid reductions in methane emissions can deliver near-term climate benefits, making methane mitigation a critical strategy for slowing warming in the coming decades [4]. The urgency of methane control is underscored by the accelerating rise in atmospheric methane concentrations, driven primarily by human activities such as fossil fuel production, waste management, and agriculture [5].

Anthropogenic methane emissions exhibit a highly skewed, heavy-tailed distribution, where a small subset of large, anomalous point sources ("super-emitters") accounts for a disproportionate share of total emissions [3, 4]. While episodic ultra-emission events from oil and gas infrastructure offer cost-effective mitigation targets [3–5], a narrow focus on these super-emitter events overlooks the significant, cumulative impact of smaller, persistent, and under-reported emissions from other sectors. For instance, recent studies indicate substantially higher methane releases from inefficient gas flaring than previously assumed [6]. Furthermore, satellite observations combined with machine learning are revealing the widespread impact of the agricultural sector, linking emissions to specific dairy farm practices and enabling new mitigation pathways [7]. Capturing this emission's "long tail" is essential for accurate global methane accounting and for formulating comprehensive climate mitigation strategies that address the full spectrum of emission sources.

[1]School of EECMS, Curtin University, Bentley WA 6102, Australia. {minh.tran2; hoa.v.nguyen; ba-ngu.vo}@curtin.edu.au.

[2]Latconnect60, 1 William St, Perth WA 6000, Australia. {aimuni; hareeshrao}@latconnect60.com.

Satellite remote sensing is instrumental in monitoring anthropogenic methane emissions [8–10]. Global monitoring relies on instruments like the TROPOspheric Monitoring Instrument (TROPOMI) aboard Sentinel-5P, which offers daily global coverage but at a coarse spatial resolution of 5.5×7 km² [11]. Conversely, high-resolution airborne hyperspectral imagers such as the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS-NG) enable detailed, targeted plume analysis [12]. This presents a trade-off between the temporal revisit rate achieved by spaceborne sensors and the fine-scale spatial detail offered by airborne platforms. The Sentinel-2 constellation, with its unique combination of global coverage, a revisit rate of $2 - 5$ days, and 20×20 m² resolution in key Short-Wave Infrared (SWIR) bands, is strategically positioned to bridge this observational gap between frequent monitoring and high-resolution detail [1]. Although not specifically designed for direct gas detection, Sentinel-2's SWIR bands exhibit sensitivity to methane absorption features, rendering them capable of detecting larger plume signatures [13].

Despite this potential, existing physics-based retrieval methods for Sentinel-2 often produce high false-positive rates, especially over heterogeneous surfaces, and require extensive manual validation [10]. Detecting smaller, more persistent leaks is particularly challenging due to low signal-to-noise ratios [1, 2, 4]. In this work, we address these challenges by introducing a deep learning approach for methane plume segmentation in Sentinel-2 imagery, with a focus on detecting small-scale emissions. Our primary contributions are:

i) A novel feature engineering approach, specifically designed to optimise the input features of satellite imagery, that combines two spectral enhancement techniques to improve the detection of methane signals.

ii) An evaluation of deep learning architectures demonstrating that a U-Net with a ResNet34 encoder, when combined with our novel feature engineering approach, overcomes the limitations of existing methods to detect single-pixel methane plumes in Sentinel-2 images.

II. RELATED WORK

Methane monitoring via satellite has advanced from physics-based retrievals, using radiative transfer and methane's spectral absorption models, to complementary deep learning models.

Multi-Sensor Methane Detection via Tiered Satellite Systems. Emerging solutions focused on tiered observation systems that combine data from multiple satellites to balance spatial resolution, temporal coverage, and detection sensitivity. For example, integrated systems leverage Sentinel-5P for daily

global methane monitoring, Sentinel-2 for high-resolution (20 m) detection, and Sentinel-3 as an intermediate sensor [4]. The use of multi-band, multi-pass (MBMP) retrieval algorithms with the SWIR bands of Sentinel-2 has proven capable of quantifying large leaks. Varon et al. [9] further demonstrated high-frequency monitoring of anomalous methane point sources with Sentinel-2, showing that operational satellites can reliably detect significant emissions by applying robust ratioing techniques. At much finer scales, very high-resolution sensors like WorldView-3 (3.7 m) have shown promise in mapping methane from specific infrastructure [8], while airborne imaging spectrometers such as AVIRIS-NG serve as critical tools for algorithm development and validation at even higher resolutions [10].

Despite their successes, traditional physics-based approaches face several challenges. Methane absorption signals in multispectral data are relatively weak and can be confounded by surface reflectance heterogeneity and varying illumination geometry [10]. As a result, simple difference or ratio methods often trigger false positives over spectrally complex landscapes. Even sophisticated regression techniques may require expert intervention to distinguish real plumes from surface artifacts [8]. Cloud edges and thin cirrus also pose a problem, frequently creating spurious "plume-like" signals in SWIR images. Consequently, these methods typically require manual post-processing and cross-verification [10], which reduces scalability and increases response times. While effective for large plumes, they often fail to reliably detect the "long tail" of smaller or more subtle emission events.

Rise of Deep Learning Models. Recently, deep learning (DL) has become a powerful alternative to classical methane detection methods. In hyperspectral imaging, Ruzicka et al. [10] introduced the STARCOP dataset, achieving significantly higher F1-scores than physics-based models by learning complex spectral patterns. For multispectral data, the S2MetNet project established a benchmark dataset, showing that DL models outperform traditional methods in quantifying methane emissions from Sentinel-2 [2]. Similarly, Vaughan et al. [1] proposed CH4Net, a U-Net-based model trained on extensive annotated Sentinel-2 imagery, setting a new standard for methane detection by effectively distinguishing real plumes from false positives.

Beyond single-model approaches, researchers are exploring techniques like data-efficient deep transfer learning to adapt models to new geographic domains with high accuracy [14] and ensemble methods that combine multiple models or multi-resolution observations to increase detection reliability [15]. While these DL approaches represent a major leap forward, accurately detecting small, persistent leaks near the detection limit of current sensors and ensuring model generalisation across diverse conditions remains an active research area. Our work builds directly on this foundation by focusing on small-plume detection through refined feature engineering to facilitate further progress.

III. DATASET AND METHODS

Our primary objective is to develop an automated and scalable system for the detection and segmentation of methane plumes from Sentinel-2 imagery, with a specific focus on identifying small plumes that are often missed by existing methods. This requires overcoming three core challenges: (i) dataset generation incorporating feature extraction techniques to support the detection of small methane plumes in multispectral data, (ii) creating a high-quality, large-scale annotated dataset of real-world plumes, and (iii) selecting and optimising a DL architecture sensitive enough for this task.

A. Feature Enhancement

A high-quality dataset is critical for training robust deep learning models. To this end, we constructed a dataset comprising real plume events, methane-free background scenes, and augmented variations. All data were sourced from Sentinel-2 imagery of an oil and gas facility in Australia, a site with well-documented CH_4 releases in 2023.

Feature Enhancement: The methane absorption characteristic in its short-wave infrared (SWIR) bands extracted from Sentinel-2 multispectral data is subtle and often obscured by background surface reflectance [10]. To overcome this, we transformed the raw satellite radiances into a multi-channel image that maximises the contrast between methane-plume and background pixels. This transformation is achieved by calculating two spectral ratios on a per-pixel basis, each generating a spatial map of ratio values - a single-channel image that highlights different aspects of the methane signal. These individual channels are then combined into a multi-channel representation. Let R_{11} and R_{12} be the top-of-atmosphere (TOA) radiances for Sentinel-2's SWIR bands 11 and 12, respectively. Band 12 (2190 nm) is sensitive to methane absorption, while nearby Band 11 (1610 nm) is largely unaffected, making it an ideal reference.

Varon Ratio $V(\cdot, \cdot)$: This ratio amplifies the methane signal by normalising the strongly methane-absorbing band R_{12} against the weakly absorbing band R_{11} [9]. It is calculated for each pixel as:

$$V(R_{12}, R_{11}) = \frac{cR_{12} - R_{11}}{R_{11}}. \qquad (1)$$

Here, the scaling factor c, computed via a least-squares fit of R_{12} against R_{11}, accounts for the combined effects of instrumental response and uniform surface spectral albedo. In pixels affected by a methane plume, atmospheric methane absorbs incoming solar radiation, suppressing R_{12} radiance, while R_{11} remains unaffected. The Varon ratio leverages this spectral contrast to highlight methane-rich pixels. The resulting output is a single-channel image in which where methane plumes become clearly visible as spatially coherent regions with notably different pixel intensities compared to the background.

Sánchez Ratio $S(\cdot)$: While effective, the Varon ratio can still be sensitive to certain surface types (water bodies) having

naturally low R_{12} to R_{11} ratios, leading to false positives. The Sánchez ratio, inspired by Sánchez-García et al. [8], uses a multi-linear regression model trained on plume-free imagery to estimate the expected radiance, $\hat{R}_{12}$, from other non-absorbing bands (Band 11 and other visible/SWIR bands), representing the radiance expected for that surface in the absence of methane. The Sánchez ratio is given by [10]:

$$S(R_{12}) = V(R_{12}, \hat{R}_{12}). \tag{2}$$

By replacing the single reference band R_{11} with a more robust, surface-adaptive prediction $\hat{R}_{12}$, this approach effectively isolates the plume signal from background spectral variations. This yields a single-channel image where, similar to the standard Varon ratio, methane plumes become easily distinguishable from normal background variation.

By stacking these two single-channel images, we formed a two-channel image where the Varon ratio (V) highlights methane signals, and the Sánchez ratio (S) reduces noise by filtering out false positives from background variability. Since many DL models require three-channel input, we duplicate V to form a pseudo-RGB image [V, S, V]. This setup reinforces the methane signal in the first and third channels, while the middle channel (S) helps suppress false positives. Not only does it satisfy deep learning input requirements but also provides a signal-enhanced representation that enables models to better learn the complex spatial and spectral patterns of methane plumes, resulting in more reliable detection than using raw bands or a single-ratio image alone.

B. Dataset Construction

Ground-Truth Generation: Ground-truth annotations for methane plumes were developed through a systematic, semi-automated labelling process tailored to emission events at an Australian site in 2023 (see Fig. 1), ensuring precise and consistent spatial annotations for supervised learning. Initial plume boundaries were established using specialised contour detection functions and predefined concentration thresholds specific to the site's primary emission sources (Vent A-D). These automated contours were then meticulously refined through manual validation to capture distinct emission patterns and address complex scenarios, such as spatially overlapping plumes, while preserving individual source information. Validation against temporal and spatial emission records minimised subjectivity and ensured consistency between spectral signatures and documented events, establishing a reliable dataset for robust model training and evaluation.

Background Data Generation: In addition to plume-containing images, we curated methane-free background scenes from Sentinel-2 observations of regions like the ground-truth facility and other oil/gas sites during periods of no known emissions. These help the model recognise "*normal*" conditions and reduce false detections. No-plume datasets are generated using a regression-based simulation to characterise natural spectral behaviour under methane-free conditions.
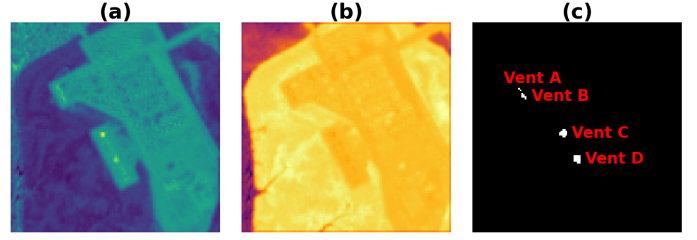

Fig. 1: Example images from the study site on 23 July 2023, showing (a) the Varon ratio, (b) the Sanchez ratio, and (c) ground-truth annotations of methane plumes (labeled as Vent A–D) highlighting emission sources.

TABLE I: Model Performance on the Validation Set.

Model	F1-Score	IoU
U-Net + MobileNetV2	0.7720	**0.7095**
U-Net + ResNet34	**0.7839**	0.6503
SegNeXt	0.7583	0.6188

An adaptive percentile-based thresholding mechanism distinguishes methane-sensitive spectral regions from background areas, establishing robust SWIR band correlations [4]. A linear regression model, trained on background pixels, predicts clean SWIR band 12 responses from SWIR band 11 inputs, preserving natural variability. Synthetic background data is enhanced with controlled Gaussian noise and cluster-based artefacts to simulate realistic observational and sensor noise conditions.

Data Augmentation and Preprocessing: To improve model generalisation, we applied a standard data augmentation pipeline, including geometric transformations (flips, rotations) and photometric adjustments (brightness, contrast). All input features were z-score normalised to ensure cross-scene consistency.

C. Deep Learning Architectures and Training

We evaluated three semantic segmentation architectures: U-Net [16] with a MobileNetV2 [17] encoder, U-Net with a ResNet34 [18] encoder, and SegNeXt [19]. The U-Net is a strong baseline for segmentation tasks, and the choice of MobileNetV2 and ResNet34 encoders allowed us to evaluate the trade-off between computational efficiency and feature extraction depth. SegNeXt, a more recent architecture, blends convolutions with transformer-based attention mechanisms to capture global context.

All models were implemented in PyTorch, leveraging the "segmentation-models-pytorch" library [20]. The encoders were initialised with ImageNet-pretrained weights to leverage transfer learning [21]. We used the Adam optimiser [22] with an exponential decay learning rate scheduler for all training runs. To handle the significant class imbalance between small plume pixels and the large background, we employed a combined loss function of Focal Loss [23] and Dice Loss [24] for the U-Net-ResNet34 and SegNeXt models. Hyperparameters for the loss function were optimised using a grid search [25].
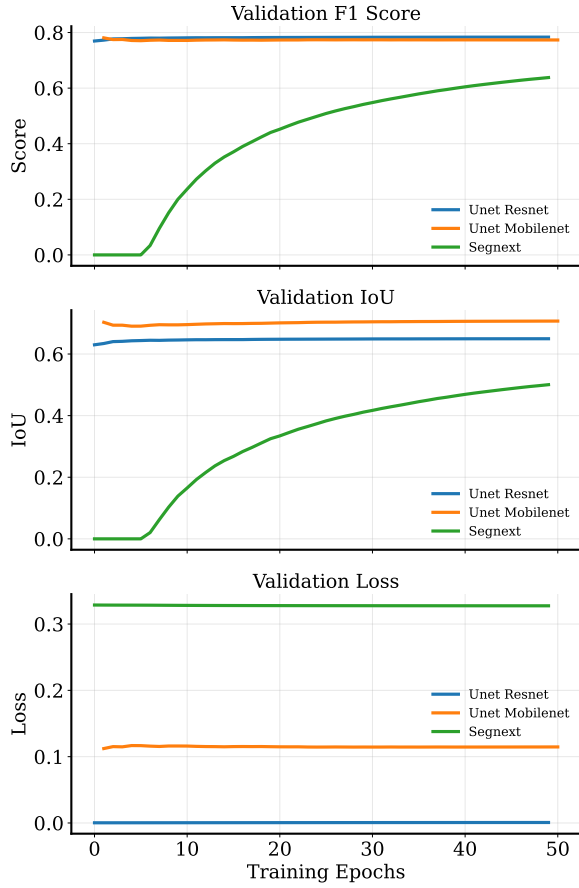
Fig. 2: Learning curves for the three architectures, showing F1-score, IoU, and loss for validation sets. The U-Net with ResNet34 encoder demonstrates the best and most stable performance.

IV. EXPERIMENTS AND RESULTS

A. Performance Metrics

The efficacy of each network architecture was evaluated using the Dice coefficient (F1-score) and IoU [26], with the loss function guiding the learning process at each step. These metrics were selected for their ability to measure overlap between predicted segmentations and ground truth, while penalising false positives—a crucial consideration for class-imbalanced datasets such as methane plume imagery [27]. In the context of methane plume detection, accurate segmentation of small, scattered plume regions demands precise boundary delineation and minimal background noise interference [26, 28]. The chosen metrics effectively capture these aspects, as higher Dice and IoU scores reflect improved edge accuracy and reduced false alarms in challenging segmentation scenarios [26].

B. Quantitative Results

We trained and evaluated the models on our curated dataset, summarising the validation set performance in Table I. The learning curves for the three architectures are presented in Fig. 2.

The U-Net with MobileNetV2 (Fig. 2) exhibited notable improvements during training, achieving a high training F1-

score of 0.9547 and training IoU of 0.9225. However, validation performance was moderate, with an F1-score of 0.7720 and an IoU of 0.7095. Although there was significant learning demonstrated, indicated by a consistent decrease in validation loss, the discrepancy between training and validation scores suggests potential overfitting and challenges in generalisation.

In contrast, the U-Net with the ResNet34 encoder (Fig. 2) demonstrated strong performance and stable learning curves, showcasing efficient learning without evident overfitting. It achieved a training F1-score of 0.9542 and a training IoU of 0.9131, closely matching MobileNetV2 in training performance but outperforming it in validation, attaining the highest validation F1-score of 0.7839. However, the validation IoU was comparatively lower at 0.6503, indicating a discrepancy between pixel-wise accuracy and overall detection capability. The residual connections within the ResNet34 encoder appear beneficial for capturing deep, relevant features effectively, enhancing segmentation robustness.

The SegNeXt model (Fig. 2) displayed steady learning progress, with training metrics at an F1-score of 0.7850 and IoU of 0.6492, and validation metrics of F1-score of 0.7583 and IoU of 0.6188. These results underscore the benefits of SegNeXt's transformer components in capturing and focusing on relevant features for plume segmentation, as noted by Guo et al. [19], its validation performance did not surpass that of the other models. The lower scores suggest limitations in handling the complex segmentation challenges presented by the dataset.

Collectively, these results indicate distinct performance contrasts across the encoder architectures. MobileNetV2 demonstrated substantial training capabilities but faced generalisation issues. The ResNet34-based U-Net emerged as the strongest performer overall, particularly in validation F1-score. SegNeXt, while showing consistent learning behaviour and reasonable segmentation accuracy, remained behind the other two architectures. Therefore, the ResNet34 encoder stands out as the most robust and suitable model for methane plume detection in this study.

C. Qualitative Results

To further assess the performance of the trained models, we present a visual comparison of their segmentation capabilities on representative test scenes (Fig. 3). These figures showcase input images, ground truth labels, predicted probability maps, binarised predictions, and difference visualisations. The "Differences" panel employs a colour-coding scheme: black for accurate non-plume predictions, red for false positives (incorrectly classified plumes), yellow for false negatives (missed plumes), and green for correctly detected plume pixels.

The U-Net with MobileNetV2 encoder (Fig. 3,(a)) demonstrates effective segmentation performance, accurately detecting plume regions with no visible false negatives. However, a few scattered false positive indicate areas where the model misclassifies background regions as plume. These suggest MobileNetV2 effectively captures plume signals but has a mild
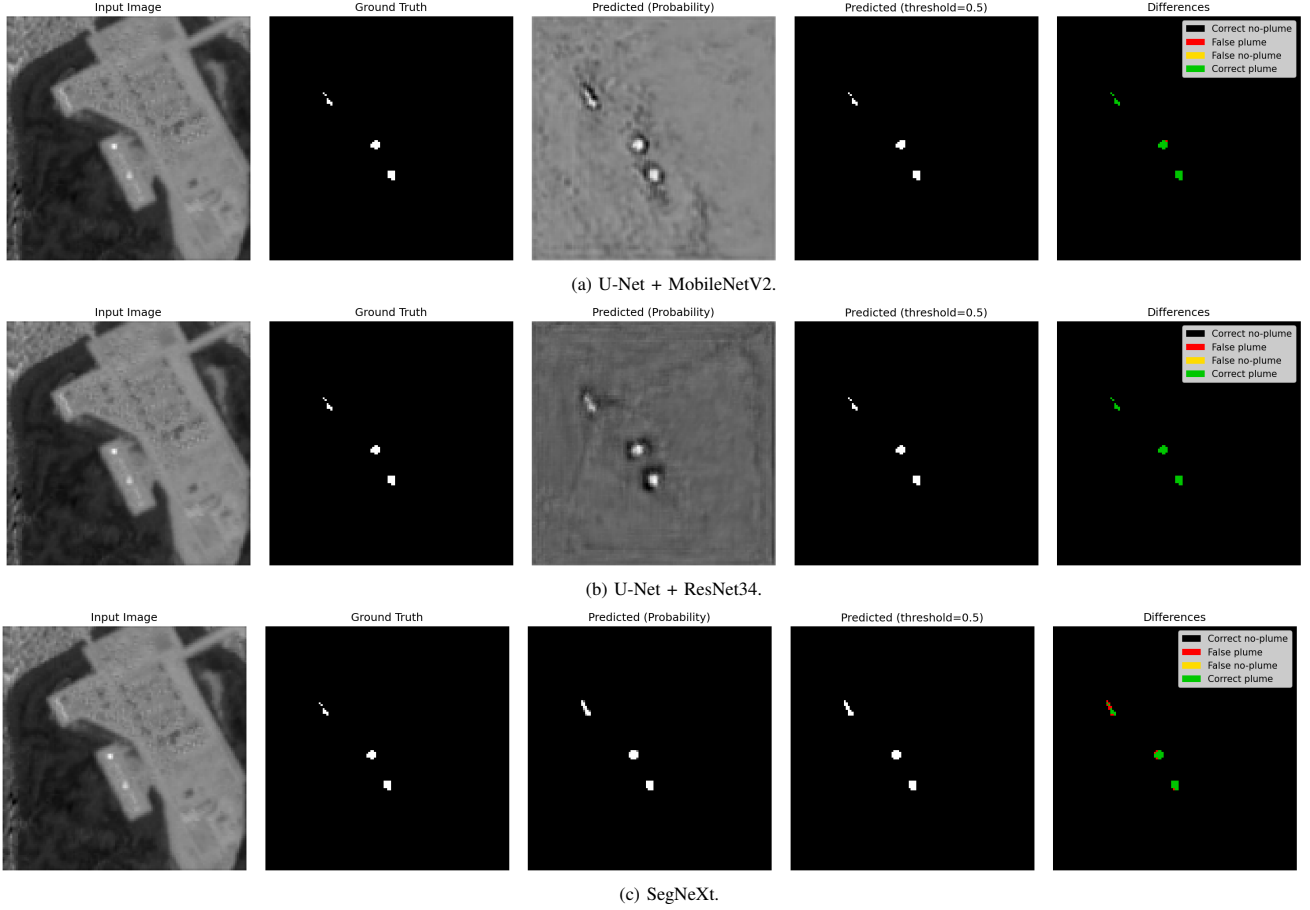
Fig. 3: Qualitative comparison of model outputs. Each panel shows (from left to right): input, ground truth, probability map, binarised prediction, and a difference map (Green: correct plume, Red: false positive, Yellow: false negative). The ResNet34-powered U-Net (b) shows the most accurate segmentation.

tendency toward over-segmentation of background areas, which aligns with its quantitative metrics.

The ResNet34 U-Net model (Fig. 3,(b)) exhibits exceptional segmentation accuracy, precisely identifying plume regions with clearly defined boundaries. Significantly, it demonstrates neither false positives nor false negatives in the provided visualisation, underscoring the model's remarkable precision and recall. These observations strongly corroborate its superior quantitative metrics, confirming ResNet34's capability in accurately differentiating plume signals from background structures without misclassifications.

The SegNeXt model (Fig. 3 (c)) detects main plume regions with few false negatives but slightly more scattered false positives than ResNet-34 U-Net. This is due to SegNeXt's global attention mechanisms, which enable the model to consider and integrate information from all parts of the image when analyzing any specific region. While this global context helps the network capture extended or subtle plume structures, it can also amplify unrelated background patterns that happen to correlate across the scene—leading to those scattered false positives. The result is excellent plume coverage with a minor overestimation, matching its strong yet slightly lower quantitative scores.

V. CONCLUSION

Our experiments show that a U-Net architecture with a ResNet34 encoder provides the best balance of feature extraction and spatial localisation, achieving superior performance over more lightweight or complex models. Most importantly, we demonstrated that this approach can reliably detect methane plumes down to a single pixel, a significant step towards a fully automated and scalable monitoring system.

For future work, we consider three directions: (i) extend the dataset and model testing to more diverse geographic regions and emission source types to improve generalizability — including multi-site cross-validation and ablation studies to quantify the contributions of the engineered spectral channels and feature pipeline; (ii) explore model interpretability techniques (e.g., saliency maps) to build trust and better integrate these tools into operational workflows for climate scientists and regulators, together with like-for-like benchmarking against recent baselines (e.g., CH4Net, S2MetNet) to situate performance; and (iii) investigate multi-temporal and ensemble approaches that leverage information across multiple satellite passes to boost sensitivity for intermittent and subtle emissions, while evaluating tailored loss functions and post-processing

to improve pixel-level delineation and reporting computational efficiency metrics (model size, memory footprint, and measured inference throughput on representative CPU/GPU) to substantiate scalability and operational readiness.

ACKNOWLEDGMENT

REFERENCES

[1] A. Vaughan, G. Mateo-García, L. Gómez-Chova, V. Růžička, L. Guanter, and I. Irakulis-Loitxate, "CH4Net: a deep learning model for monitoring methane super-emitters with sentinel-2 imagery," *EGUsphere*, vol. 2023, pp. 1–17, 2023.

[2] A. Radman, M. Mahdianpari, D. J. Varon, and F. Mohammadimanesh, "S2MetNet: A novel dataset and deep learning benchmark for methane point source quantification using Sentinel-2 satellite imagery," *Remote Sensing of Environment*, vol. 295, p. 113708, 2023.

[3] T. Lauvaux, C. Giron, M. Mazzolini, A. d'Aspremont, R. Duren, D. Cusworth, D. Shindell, and P. Ciais, "Global assessment of oil and gas methane ultra-emitters," *Science*, vol. 375, no. 6580, pp. 557–561, 2022.

[4] S. Pandey, M. van Nistelrooij, J. D. Maasakkers, P. Sutar, S. Houweling, D. J. Varon, P. Tol, D. Gains, J. Worden, and I. Aben, "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*, vol. 296, p. 113716, 2023.

[5] R. A. Alvarez, D. Zavala-Araiza, D. R. Lyon, D. T. Allen, Z. R. Barkley, A. R. Brandt, K. J. Davis, S. C. Herndon, D. J. Jacob, A. Karion *et al.*, "Assessment of methane emissions from the US oil and gas supply chain," *Science*, vol. 361, pp. 186–188, 2018.

[6] G. Plant, E. A. Kort, A. R. Brandt, Y. Chen, G. Fordice, A. M. Gorchov Negron, S. Schwietzke, M. Smith, and D. Zavala-Araiza, "Inefficient and unlit natural gas flares both emit large quantities of methane," *Science*, vol. 377, no. 6614, pp. 1566–1571, 2022.

[7] H. Bi and S. Neethirajan, "Mapping methane—the impact of dairy farm practices on emissions through satellite data and machine learning," *Climate*, vol. 12, no. 12, p. 223, 2024.

[8] E. Sánchez-García, J. Gorroño, I. Irakulis-Loitxate, D. J. Varon, and L. Guanter, "Mapping methane plumes at very high spatial resolution with the WorldView-3 satellite," *Atmospheric Measurement Techniques Discussions*, pp. 1–26, 2021.

[9] 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 Discussions*, vol. 2020, pp. 1–21, 2020.

[10] 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*, vol. 13, no. 1, p. 19999, 2023.

[11] D. J. Varon, D. J. Jacob, B. Hmiel, R. Gautam, D. R. Lyon, M. Omara, M. Sulprizio, L. Shen, D. Pendergrass, H. Nesser *et al.*, "Continuous weekly monitoring of methane emissions from the permian basin by inversion of tropomi satellite observations," *Atmospheric Chemistry and Physics Discussions*, vol. 2022, pp. 1–26, 2022.

[12] S. Jongaramrungruang, A. K. Thorpe, G. Matheou, and C. Frankenberg, "MethaNet–An AI-driven approach to quantifying methane point-source emission from high-resolution 2-D plume imagery," *Remote Sensing of Environment*, vol. 269, p. 112809, 2022.

[13] A. Groshenry, C. Giron, T. Lauvaux, A. d'Aspremont, and T. Ehret, "Detecting methane plumes using PRISMA: Deep learning model and data augmentation," *NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning.*, 2022.

[14] S. Zhao, Y. Zhang, S. Zhao, X. Wang, and D. J. Varon, "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*, 2025.

[15] A. K. Ayasse, D. H. Cusworth, K. Howell, K. O'Neill, B. M. Conrad, M. R. Johnson, J. W. Heckler, G. P. Asner, and R. M. Duren, "Probability of detection and multi-sensor persistence of methane emissions from coincident airborne and satellite observations," *Environmental Science & Technology*, vol. 58, pp. 21 536 – 21 544, 2024.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[17] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2015.

[19] M.-H. Guo, C.-Z. Lu, Q. Hou, Z. Liu, M.-M. Cheng, and S.-M. Hu, "Segnext: Rethinking convolutional attention design for semantic segmentation," *Advances in neural information processing systems*, vol. 35, pp. 1140–1156, 2022.

[20] P. Iakubovskii, "Segmentation models pytorch," https://github.com/qubvel/segmentation_models.pytorch, 2019.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, pp. 84 – 90, 2012.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[23] T.-Y. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2999–3007, 2017.

[24] F. Milletarì, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, 2016.

[25] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, 2012.

[26] L. Xu, X. Ji, J. Xiao, Q. Gao, L. Tang, B. Ou, and C. Li, "Infrared imaging detection for gas leakage based on residual-enhanced nnunetv2: Taking methane leakage from a natural gas pipeline as a case study," *SSRN Electronic Journal*, 2025.

[27] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, 2022.

[28] C. Chen, M. Fan, Z. Wang, M. Liang, J. Tao, and L. Chen, "Mpsunet: A deep learning-based segmentation framework for methane plume detection with space-based hyperspectral and multispectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–15, 2025.