A Self-Supervised Learning Framework for Methane Emission Detection Using Sentinel-2

Mohammad Marjani[1], Masoud Mahdianpari[1, 2], Eric W Gill[1], Fariba Mohammadimanesh[3]

[1] Department of Electrical and Computer Engineering, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada – mmarjani@mun.ca; m.mahdianpari@mun.ca; ewgill@mun.ca
[2] C-Core, St. John's, NL A1B 3X5, Canada - masoud.mahdianpari@c-core.ca
[3] Canada Centre for Remote Sensing, Natural Resources Canada, 580 Booth Street, Ottawa, ON, Canada - fariba.mohammadimanesh@nrcan-rncan.gc.ca

Keywords: Methane (CH4) detection, Self-Supervised Learning (SSL), Deep Learning, Simple Framework for Contrastive Learning of Visual Representations (SimCLR), Remote Sensing, Sentinel-2.

Abstract

Methane (CH4) is a major greenhouse gas; however, large-scale monitoring remains challenging due to the high costs and spatial limitations of ground-based and airborne observations. In contrast, Sentinel-2 shortwave infrared (SWIR)–based plume detection is hindered by its coarse spectral resolution, surface artifacts, and limited real-world annotations. This study proposes a self-supervised learning (SSL) framework based on the Simple Framework for Contrastive Learning of Visual Representations (SimCLR) to learn transferable CH4 plume representations from unlabeled Sentinel-2 data. A real-world dataset of 456 Sentinel-2 image tiles was manually annotated using the multi-band–multi-pass (MBMP) approach and utilized to evaluate six encoder backbones. Across five labeled-data portions ranging from 20% to 100%, SimCLR pretraining improved plume segmentation compared to ImageNet-only initialization. In the full-data scenario, MobileNet achieved an F1-score of 0.90 with an Intersection over Union (IoU) of 0.80, while Shifted Window Transformer (SwinT) reached an F1-score of 0.85 with an IoU of 0.75. The benefit of self-supervised pretraining was most evident with limited labeled data, where ImageNet-only models degraded substantially, while SimCLR-pretrained encoders achieved higher accuracy. Moreover, the Integrated Mass Enhancement (IME) method was employed for quantifying the emission flux rate. MobileNet provided the strongest agreement with reference emission estimates, achieving an RMSE of 1690 kg/h. Finally, the results demonstrate that SimCLR-based SSL substantially enhances CH4 plume detection from Sentinel-2 imagery and supports more reliable emission quantification for large-scale CH4 monitoring.

1. Introduction

Since the Industrial Revolution, global warming has been primarily driven by radiative forcing from greenhouse gases (GHGs), with methane (CH4) being a significant contributor (Change, 2013). CH4 has a global warming potential 27–30 times higher than carbon dioxide over a 100-year horizon (Etminan et al., 2016) and an atmospheric lifetime of roughly one decade (Guanter, 2021). A 50% reduction in anthropogenic CH4 emissions by 2050 could prevent up to 0.25 °C of warming by mid-century and as much as 0.5 °C by 2100 (Rouet-Leduc et al., 2023), which highlights the importance of rapid CH4 mitigation strategies.

Conventional ground-based and airborne monitoring methods, such as in-situ sensors, patrol surveys, drones, and aircraft, involve high operational costs and mainly capture localized emissions (Rouet-Leduc et al., 2023). These limitations restrict their suitability for large-scale CH4 monitoring, discovery of unknown emission sources, and development of regional or global emission inventories (Mohammadimanesh et al., 2025). Satellite remote sensing has therefore become a key tool for CH4 observation, particularly for detecting plumes from point sources (Marjani et al., 2024). CH4 detection relies on shortwave infrared (SWIR) observations, where absorption features occur in the 1600–2500 nm spectral range (Marjani et al., 2025). Both hyperspectral and multispectral satellite data have been used for plume detection within these SWIR bands (Guanter, 2021; Marjani et al., 2024; Radman et al., 2023; Varon et al., 2021). Hyperspectral sensors offer superior spectral resolution but often lack the spatial resolution and revisit frequency required for continuous monitoring of localized emission sources.

The Sentinel-2 mission offers a unique balance of global coverage, high spatial resolution, and a five-day revisit cycle (Radman et al., 2023). However, its relatively coarse SWIR bands limit sensitivity to CH4 absorption, which results in detection thresholds on the order of 1800–2000 kg h^{-1} under favorable conditions (Irakulis-Loitxate et al., 2022; Varon et al., 2021). Several CH4 retrieval techniques based on Sentinel-2 band ratios and temporal differencing have been proposed, including single-band–multi-pass (SBMP), multi-band–single-pass (MBSP), and multi-band–multi-pass approaches (MBMP) (Varon et al., 2021). These methods have been adopted in multiple studies (Gorroño et al., 2023; Radman et al., 2023), followed by plume boundary extraction through manual or statistical thresholding. Manual CH4 plume boundary extraction requires substantial effort, while threshold-based methods often misclassify surface artifacts as CH4 emissions. In addition, approaches that rely on multi-pass observations are not applicable at sites with limited temporal coverage.

Recent studies have explored deep learning methods to overcome these challenges, although applications to Sentinel-2-based CH4 plume detection remain limited. Existing approaches rely heavily on synthetic datasets to compensate for the scarcity of annotated real-world plumes. While synthetic data provide controlled plume scenarios, they fail to fully represent background heterogeneity, atmospheric variability, and surface artifacts. As a result, models trained on synthetic or region-specific datasets often show reduced transferability to unseen locations or emission sources. Furthermore, most available annotated datasets contain only a small number of plume instances, which constrains the robustness of supervised learning approaches. Threshold-dependent segmentation methods also introduce uncertainty in plume boundary detection and frequently produce false positives. This study addresses these limitations through a self-supervised learning (SSL) framework based on the Simple Framework for Contrastive Learning of Visual Representations (SimCLR). The proposed approach learns spatial–spectral representations of CH4 plumes from unlabeled Sentinel-2 imagery, thereby reducing reliance on synthetic data and the need for extensive manual annotation. Contrastive learning enables feature extraction that remains robust across diverse backgrounds and environmental conditions. The main objectives of this study are to:

(1) develop an SSL framework based on SimCLR for learning CH4 plume representations from Sentinel-2 data without labeled pretraining data;

(2) construct a real-world CH4 plume dataset through manual annotation across multiple point sources and environments;

(3) compare different encoder architectures, including convolutional and transformer-based models, for plume detection;

(4) evaluate model performance under limited labeled data and unseen emission scenarios; and

(5) estimate CH4 emission flux rates using the Integrated Mass Enhancement (IME) method applied to detected plume boundaries.

2. Material and Method

2.1 Satellite Data and Preparation

This study utilizes Sentinel-2 multispectral imagery, which consists of two identical satellites, Sentinel-2A and Sentinel-2B. Although Sentinel-2 was not designed for CH4 monitoring, its SWIR bands (Band 11 (1560–1660 nm) and Band 12 (2090–2290 nm)) exhibit sensitivity to CH4 absorption features (Radman et al., 2023). These bands, available at 20 m spatial resolution, support detection of strong CH4 point-source emissions under favorable atmospheric and surface conditions. Despite the coarse spectral resolution of these SWIR bands, Sentinel-2 has demonstrated strong capability for large-scale identification of significant CH4 emission sources.

A real-world CH4 plume dataset was collected from Sentinel-2 imagery acquired between 2017 and 2023. The dataset covers a wide range of emission sources, including oil and gas facilities, landfills, and coal mines, as well as diverse environmental and surface conditions. In total, 456 image tiles were collected, each with a spatial resolution of 20 m and a fixed size of 256 × 256 pixels. CH4 plume detection was formulated as a binary segmentation task, where pixels were labeled as either plume or non-plume.

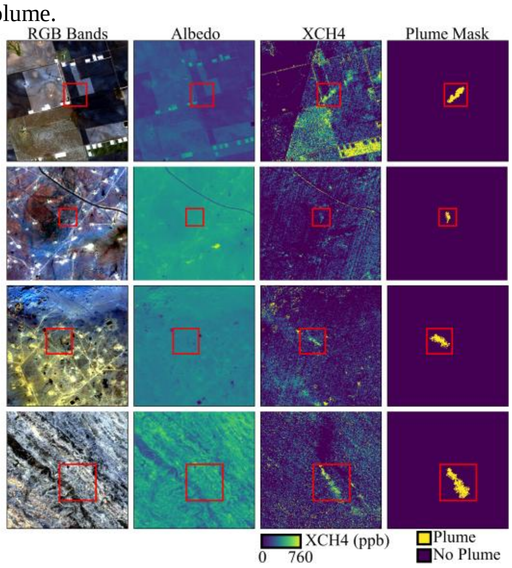

Figure 1. Representative samples of the Sentinel-2 CH4 plume dataset across multiple locations. Red squares highlight the plume regions.

Manual annotation followed a physically guided procedure based on the MBMP enhancement method (Varon et al., 2021). For each site, a clear-sky reference image was selected, and MBMP concentration maps were generated. CH4 plumes visible in MBMP concentration maps and spatially consistent with wind direction were manually delineated using polygon boundaries. Regions correlated with surface artifacts, albedo variations, or visible features in RGB composites were excluded to reduce false labeling. Plumes detectable directly from true-color imagery were also removed to avoid bias toward visually obvious cases. Examples of annotated plume masks and corresponding enhancement maps are shown in Figure 1.

2.2 SimCLR-Based Self-Supervised Learning (SSL)

SSL is a representation learning approach that derives training signals directly from unlabeled data. Instead of relying on manual annotations, SSL defines a proxy objective that encourages a model to learn transferable features from a large dataset of images. This structure is well-suited for satellite remote sensing, where imagery is abundant but pixel-level labels are scarce and expensive (Wang et al., 2022). In this study, SSL is employed to learn general spatial–spectral representations from Sentinel-2 data, followed by supervised fine-tuning for CH4 plume segmentation.

Contrastive learning is a widely used SSL strategy that learns by comparing samples. The central idea is to pull representations of similar samples closer in feature space while pushing representations of dissimilar samples farther apart. Similarity is defined by creating different views of the same image through data augmentation. This design promotes invariance to nuisance factors commonly found in satellite imagery, including illumination changes, background heterogeneity, and moderate atmospheric variability.

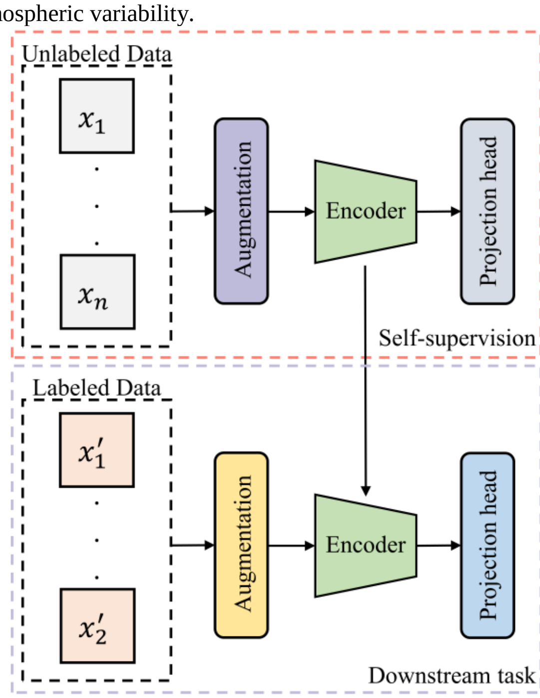

Figure 2. The general overview of the SSL framework.

This study adopts SimCLR (Chen et al., 2020), which forms a positive pair by generating two independently augmented views from the same Sentinel-2 image, while augmented views from other images in the same mini-batch serve as negative samples.

The SimCLR pipeline contains four main elements: (1) a stochastic augmentation module that creates diverse but consistent views of the same scene, (2) a shared encoder that extracts feature representations from each view, (3) a projection head (a small neural network) used only during SSL pretraining to improve representation quality, and (4) a contrastive objective that promotes agreement between positive pairs and separation from negatives. An overview of the SimCLR approach is shown in Figure 2.

2.2.1 Data Augmentation Techniques: Data augmentation plays a central role in contrastive SSL by producing diverse yet semantically consistent views of each input image. Within the SimCLR framework, two augmented views are generated from every Sentinel-2 image to form positive pairs for contrastive training. These transformations encourage the encoder to learn invariant spatial–spectral representations that remain stable under realistic scene variability.

The augmentation strategy primarily focuses on spatial transformations that alter the geometric structure while preserving plume semantics. In this study, random zoom, rotation, horizontal and vertical flipping, translation, and cutout operations are applied independently and stochastically to each view.

2.2.2 Encoders: In this study, both convolutional and transformer-based encoders are evaluated within the SimCLR framework, including MobileNet (Howard et al., 2017), Inception-v3 (Szegedy et al., 2016), DenseNet-121 (Huang et al., 2017), Xception (Chollet, 2017), Vision Transformer (ViT) (Dosovitskiy et al., 2021), and Shifted Window Transformer (SwinT) (Liu et al., 2021). For all encoders, only the feature extraction backbone is used, and task-specific classification heads are excluded to maintain a consistent and fair comparison across models.

2.2.3 Projection Head: After SSL pretraining, the projection head is removed and the pretrained encoder is transferred to the supervised CH4 plume segmentation task using limited labeled data. This transfer reduces dependence on synthetic datasets and supports better generalization to unseen emission sources and environmental settings. In this study, the projection head is implemented as a shallow multilayer perceptron composed of three fully connected layers with 256 units each and Rectified Linear Unit (ReLU) activation functions.

2.2.4 Normalized Temperature-Scaled Cross-Entropy (NT-Xent) Loss Function: Within the SimCLR framework, representation learning is driven by a contrastive loss function that promotes similarity between different augmented views of the same Sentinel-2 image while enforcing dissimilarity between views originating from other images. For a given positive pair (i, j), where both samples come from the same image, the normalized temperature-scaled cross-entropy (NT-Xent) loss is defined using Equation (1):

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} \log\left(\frac{\exp(sim(Z_i, Z_{i'})/\tau)}{\sum_{j=1}^{2N} 1_{|j\neq i|} exp\,(sim(Z_j, Z_{j'})/\tau)}\right) \quad (1)$$

where N represents the number of images in a batch, Z_i and $Z_{i'}$ are the feature representations of the two augmented views of the same image, while Z_j and $Z_{j'}$ represents the features from all other images in the batch. The function sim denotes the cosine similarity between the representations. Moreover, the temperature parameter τ serves as a scaling factor for the similarity scores, influencing the distribution of the logits used in the SoftMax function.

2.3 Fine-Tuning Strategy

Fine-tuning adapts the self-supervised encoder pretrained within the SimCLR framework to the supervised CH4 plume segmentation task using labeled Sentinel-2 imagery. During self-supervised pretraining, the encoder learns robust spatial–spectral representations by associating different augmented views of the same image, while a projection head supports contrastive learning. For downstream segmentation, the projection head is removed and replaced with a task-specific decoder.

As illustrated in Figure 3, a decoder composed of transpose convolution (ConvT) layers and batch normalization (BN) is appended to the encoder to perform pixel-wise plume classification. This decoder architecture remains fixed across all encoder backbones to allow consistent comparison. The combined encoder–decoder model produces binary segmentation maps that classify each pixel as plume or non-plume.

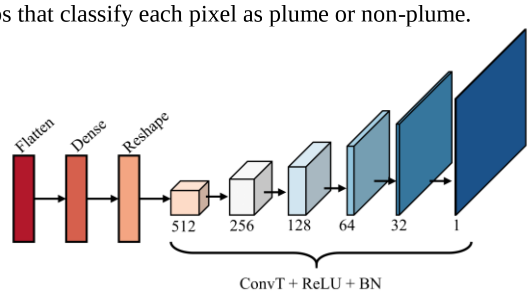

Figure 3. The architecture of the decoder used in this study.

Two fine-tuning scenarios are investigated, both following a two-stage training strategy designed to preserve learned representations while stabilizing supervised optimization. An overview of these strategies is provided in Figure 4. In the first scenario, the encoder is initialized using weights obtained from SimCLR-based self-supervised pretraining. During the initial warm-up phase, the encoder remains frozen for 10 epochs while the decoder is trained independently. This stage allows the decoder to learn spatial mappings without altering the pretrained encoder features. In the subsequent stage, the encoder and decoder are trained jointly using supervised labels, with a lower learning rate applied to the encoder to retain its pretrained representations.

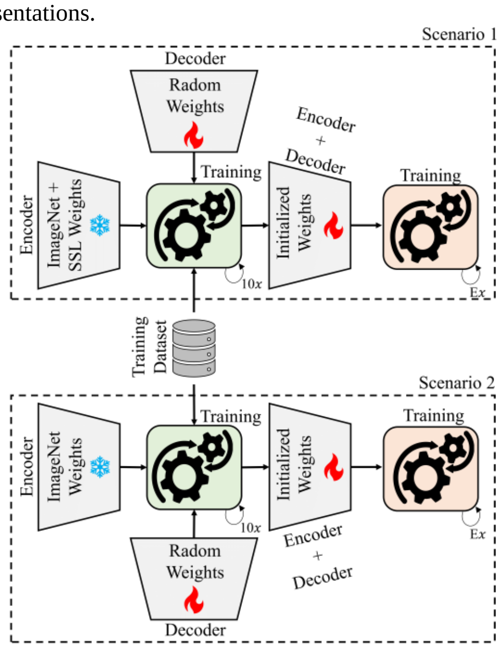

Figure 4. Schematic of the two-stage fine-tuning strategies adopted in this study. In Scenario 1 (top), the encoder is initialized using SimCLR-based self-supervised and ImageNet-pretrained weights. The decoder is randomly initialized and trained for the first 10 epochs while the encoder remains frozen (snowflake icon), followed by joint fine-tuning of both components (flame icon) for Ex epochs. In Scenario 2 (bottom), the encoder is initialized with ImageNet-pretrained weights only and trained using the same two-stage protocol.

In the second scenario, the encoder is initialized using ImageNet-pretrained weights only, representing a conventional transfer learning baseline. The same two-stage training protocol is applied: the encoder is frozen during the first 10 epochs while the decoder is trained, followed by joint fine-tuning of both components. This setup enables a direct comparison between self-supervised pretraining and standard supervised initialization under identical training conditions.

2.4 CH4 Characterization

CH4 emission rates (Qs) are estimated for each detected plume during the final processing stage. Accurate plume detection is required before quantification and is achieved using the proposed framework. In this study, Qs are calculated using the IME method, which has demonstrated robust performance in satellite-based CH4 quantification across multiple studies (Marjani et al., 2024, 2025; Radman et al., 2023; Varon et al., 2021). Moreover, uncertainty in emissions is evaluated by considering three primary contributors: wind speed variability, model assumptions of the IME, such as wind coefficients, and retrieval-related uncertainty. Wind speed uncertainty represents the dominant source of error due to the coarse resolution of meteorological data relative to plume-scale dynamics (Mohammadimanesh et al., 2025). Retrieval uncertainty is assessed based on variability in CH4 enhancement within the plume mask and surrounding background regions. These factors provide an overall assessment of uncertainty associated with the estimated CH4 Qs.

3. Results

3.1 Detection Results

The quantitative performance of the models is evaluated using Recall, Precision, Intersection over Union (IoU), and F1-score. Figures 5 and 6 compare model performance under SimCLR-based self-supervised pretraining and ImageNet-only initialization, respectively, across five training data portions (20%, 40%, 60%, 80%, and 100%). Across all metrics and data portions, models fine-tuned with SimCLR-pretrained encoders consistently outperform those initialized solely with ImageNet weights, which highlights the effectiveness of self-supervised contrastive pretraining for CH4 plume segmentation.
With SimCLR pretraining (see Figure 5), MobileNet, SwinT, and Xception achieve the strongest and most stable performance across most scenarios. At full data availability, MobileNet achieves a Recall of 0.92 and an IoU of 0.80, with balanced Precision and an F1-score of 0.90, indicating accurate plume detection with limited false alarms. SwinT exhibits comparable recall of 0.92 with slightly lower spatial overlap, IoU of 0.75, and F1-score of 0.85, whereas Xception maintains competitive performance with an F1-score of 0.81. Under reduced training data, performance decreases for all models; however, SimCLR-pretrained encoders retain stronger generalization. For example, with 20% data availability, SwinT and Xception maintain achieved IoU values of 0.10 and 0.17, respectively, and F1-scores of 0.18 and 0.27. In contrast, MobileNet exhibits a very high Recall of 0.94 at the expense of Precision (0.03), reflecting an increase in false positive detections.

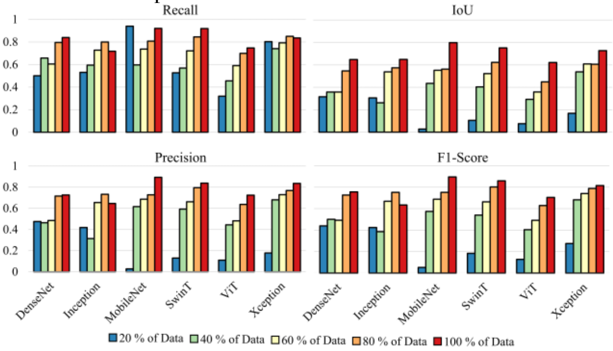

Figure 5. Quantitative performance of fine-tuned models initialized with SimCLR-pretrained encoders across five training data portions (20%, 40%, 60%, 80%, and 100%).

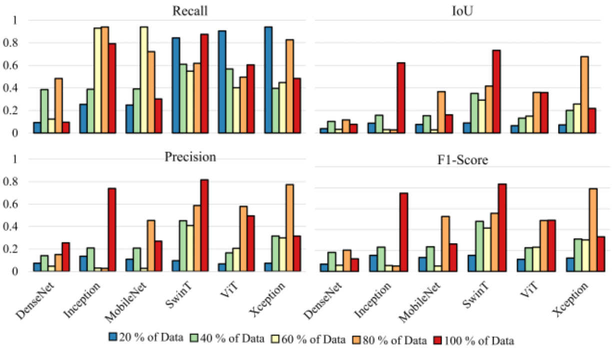

Figure 6. Quantitative performance of fine-tuned models initialized with ImageNet-pretrained weights only, without SimCLR pretraining, across five training data portions (20%, 40%, 60%, 80%, and 100%).

DenseNet and ViT also benefit from SimCLR pretraining but remain less competitive than the top-performing models. At 100% data, DenseNet achieves an F1-score of 0.75, while ViT reaches 0.70, showing lower spatial overlap than MobileNet, SwinT, and Xception. Inception displays variable behavior, with Recall increasing from 0.53 at 20% data to 0.72 at full data availability, but its Precision and IoU fluctuate considerably, which limits its overall segmentation performance.
In contrast, models trained without SimCLR pretraining (see Figure 6) experience substantial performance degradation across all metrics. At full data availability, MobileNet's F1-score drops to 0.26, while DenseNet performs poorly with an F1-score of only 0.12. ViT also struggles without SimCLR, achieving an IoU of 0.36 and an F1-score of 0.49, despite moderate Recall. SwinT remains the most resilient architecture in the ImageNet-only scenario, reaching an F1-score of 0.83, although this performance consistently falls short of that achieved with SimCLR pretraining.
Performance gaps between SimCLR-based and ImageNet-only initialization widen further under limited data conditions. At 20% data availability, models without SimCLR exhibit very low IoU values (approximately 0.07–0.09) and F1-scores around 0.15 or lower, accompanied by poor Precision, which indicates frequent misclassification of background regions as CH4 plumes. These results demonstrate that supervised-only training fails to represent robust plume features under data scarcity, whereas

SimCLR pretraining substantially improves segmentation accuracy, spatial consistency, and generalization.

In addition to the quantitative results, Figure 7 presents qualitative CH4 plume detections for six representative Sentinel-2 scenes acquired on different dates, comparing the outputs of all evaluated encoders with those initialized using SimCLR weights. Overall, models pretrained with SimCLR accurately detected plume geometry, spatial extent, and transport direction across diverse plume shapes and background conditions. MobileNet, Xception, and SwinT produce the most coherent and spatially continuous plume structures, closely matching the patterns visible in the concentration maps. These models demonstrate strong suppression of background noise and reduced sensitivity to surface artifacts, particularly in scenes with diffuse or fragmented plumes. DenseNet and Inception recover the main plume cores but exhibit partial plume fragmentation, losing fine-scale structure. ViT detects plume regions consistently but tends to produce smoother and less sharply defined boundaries, especially for narrow or low-contrast plumes. Across all cases, the qualitative results align well with the quantitative findings, confirming that SimCLR-pretrained models, especially MobileNet, Xception, and SwinT, provide more robust and physically consistent CH4 plume detection across varying emission strengths and environmental conditions.

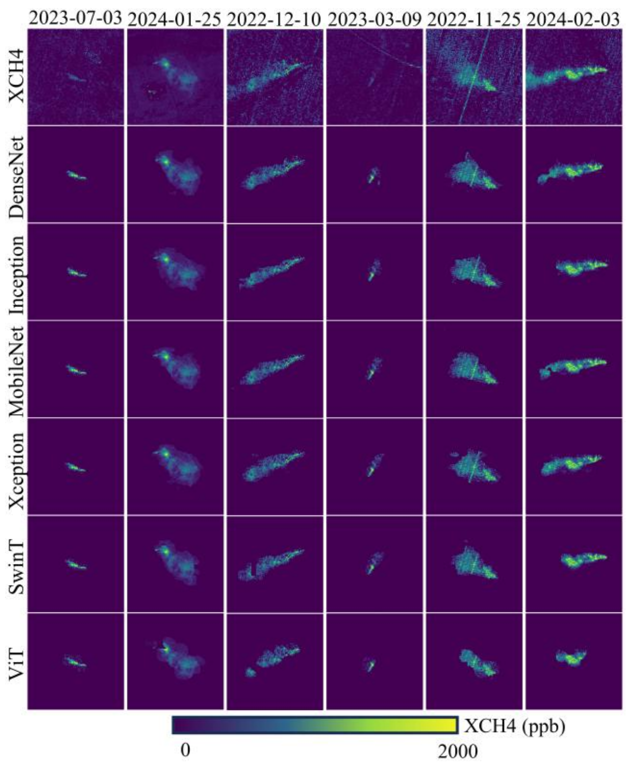

Figure 7. Qualitative comparison of CH4 plume detection for six representative Sentinel-2 scenes acquired on different dates.

3.2 Quantification Results

Following the binary plume detection, Figure 8 presents the CH4 emission flux rate quantification results obtained using the IME method, where plume masks predicted by each model are used as inputs. The figure compares predicted Qs against target values derived from ground-truth plume masks, together with a fitted regression line and the ideal 1:1 correspondence. The models exhibit varying levels of agreement with the reference Qs, indicating that segmentation quality directly influences IME-based quantification accuracy.

Among all models, MobileNet achieves the strongest agreement with the target Qs, yielding the highest coefficient of determination with R-squared (R^2) of 0.93 and the lowest error root mean square error (RMSE) of 1690 kg/h. Its regression slope remains close to unity, which indicates minimal bias and stable performance across the emission range. SwinT and DenseNet also demonstrate strong quantification capabilities, with R^2 values of 0.91 and 0.87, and RMSE values of 1,977 kg/h and 2,270 kg/h, respectively. Both models show regression slopes close to 1, indicating largely proportional estimates with slight underestimation at higher Qs.

In contrast, Inception, Xception, and ViT show reduced quantification performance. Inception and ViT yield lower R^2 values of 0.79 and higher RMSE values of 2818 kg/h and 2809 kg/h, respectively, indicating increased scatter and weaker agreement with the reference emissions. Their regression slopes below unity suggest systematic underestimation of larger Qs. Xception achieves a regression slope close to 1 but exhibits larger dispersion, with an R^2 of 0.85 and RMSE of 2596 kg/h, reflecting reduced stability in plume area estimation.

Across all models, strong agreement is observed for low to moderate Qs, as indicated by the dense clustering of samples near the origin. However, several outliers appear at higher emission levels (above approximately 20,000 kg/h), where discrepancies between predicted and target values increase substantially. These outliers correspond to cases where predicted plume masks either miss diffuse plume regions or overestimate plume extent, which directly propagates into IME-based emission errors.

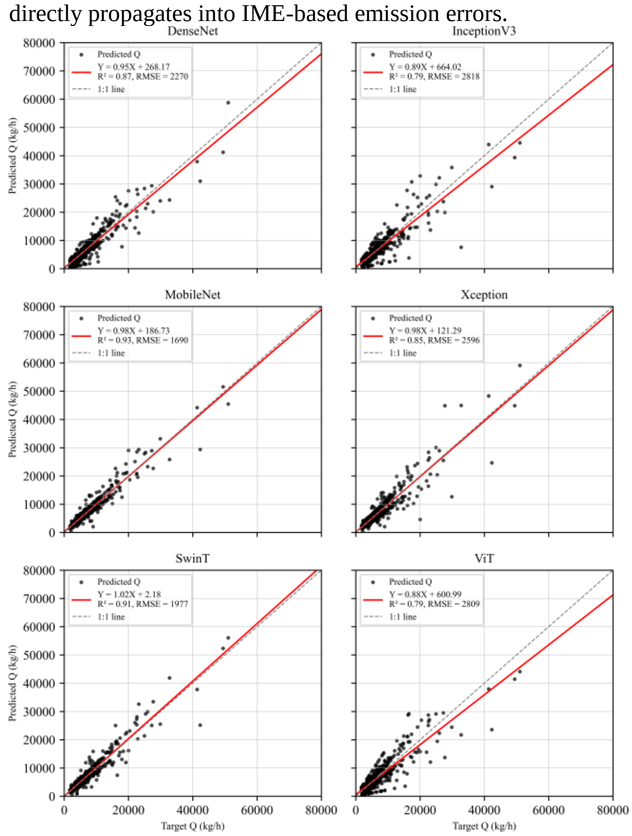

Figure 8. Quantification of CH4 Qs using the IME method based on plume masks predicted by each model. Scatter plots compare predicted Qs against target values derived from ground-truth plume masks. The solid red line indicates the linear regression fit, while the dashed line represents the ideal 1:1 correspondence. The regression equation, R^2, and RMSE are

reported for each model to summarize quantification performance.

3.3 Uncertainty Estimation

Figure 9 presents the uncertainty bounds associated with CH4 emission flux rate estimates derived using the IME method for all models. For each detected plume, predicted Qs are sorted and displayed with their corresponding uncertainty ranges, shown as shaded regions above and below the predicted values. Across all models, uncertainty increases with higher Qs, which reflects the combined influence of plume mask variability, CH4 retrieval noise, and wind speed estimation errors, particularly for large plumes. Among the evaluated models, SwinT and MobileNet exhibit the highest mean uncertainties, approximately 2783 kg/h and 2753 kg/h, respectively, consistent with their wider spread of predicted Qs. Observations also show relatively high uncertainty (2731 kg/h), indicating less stable emission estimates. In contrast, ViT yields the lowest mean uncertainty (2622 kg/h), followed by DenseNet (2672 kg/h) and Xception (2673 kg/h), suggesting more consistent quantification performance. Uncertainty bounds remain relatively narrow for low-emission plumes, indicating higher confidence in small-source estimates, while larger plumes exhibit substantially wider uncertainty ranges, highlighting the increased sensitivity of IME-based quantification to wind speed errors and plume boundary inaccuracies at high emission levels.

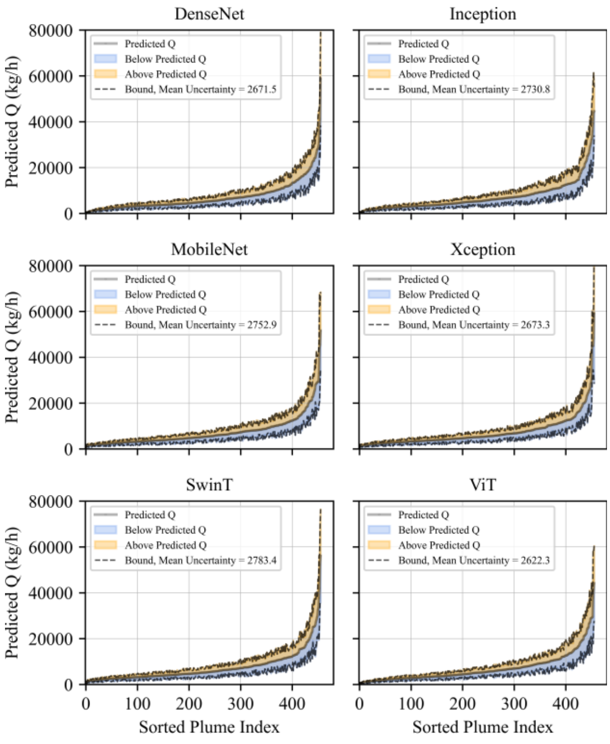

Figure 9. Uncertainty bounds of CH4 Q estimates obtained using the IME method for all evaluated models.

4. Discussion

This study demonstrates that SimCLR-based SSL provides a strong foundation for CH4 plume segmentation from Sentinel-2 imagery, particularly when labeled data are limited. The quantitative results show consistent gains for SimCLR-pretrained encoders over ImageNet-only initialization across Recall, Precision, IoU, and F1-score, which supports the primary motivation behind SSL in remote sensing (using abundant unlabeled imagery when pixel-level annotations are scarce and costly to obtain). These improvements also address key limitations reported in prior Sentinel-2 CH4 detection studies, which rely on synthetic plumes and heuristic thresholding, often reducing transferability and increasing false detections in real scenes (Radman et al., 2023; Rouet-Leduc et al., 2023; Vaughan et al., 2023; Zhao et al., 2025). The qualitative examples reinforce this outcome, where SimCLR-pretrained models represent plume geometry, spatial extent, and transport direction across diverse backgrounds, supporting the claim that contrastive representation learning can reduce sensitivity to surface artifacts and scene variability.

Differences across encoder architectures highlight how model design affects plume segmentation. Under SimCLR pretraining, MobileNet, SwinT, and Xception show the strongest performance and stability, including in reduced-label settings. MobileNet achieves high spatial overlap and balanced segmentation performance at full data availability, while SwinT maintains robust results and strong generalization, and Xception remains competitive. In contrast, DenseNet and Inception achieve moderate performance with less stability across splits, and ViT remains comparatively weaker, especially without SimCLR initialization. These trends align with the qualitative patterns in Figure 7, where MobileNet, SwinT, and Xception produce more coherent and spatially continuous plume masks. In contrast, DenseNet and Inception recover plume cores but exhibit more fragmentation, while ViT tends to generate smoother boundaries that can miss narrow or low-contrast plume structures.

The downstream CH4 characterization results indicate that segmentation quality directly affects the reliability of IME-based emission estimation. Results show that models producing more accurate plume masks yield higher agreement with reference Qs derived from ground-truth masks. MobileNet provides the strongest quantification performance, followed by SwinT and DenseNet, confirming that physically meaningful plume delineation improves IME-based source-rate retrieval (Irakulis-Loitxate et al., n.d., 2021; Varon et al., 2020).

Uncertainty patterns further support the dependence of quantification reliability on the quality of plume detection and external drivers. Across all models, uncertainty increases as Qs rise, reflecting the compounded effects of plume boundary variability, retrieval noise, and wind-related error sources. This behavior is consistent with known limitations in CH4 quantification when wind information and plume-scale dynamics are mismatched, particularly because meteorological products remain coarse relative to plume structure (Marjani et al., 2024; Sánchez-García et al., 2021; Varon et al., 2021).

5. Conclusion

This study developed a SimCLR-based SSL framework for CH4 plume segmentation from Sentinel-2 imagery and evaluated its ability to reduce reliance on large labeled datasets and synthetic plume generation. A real-world dataset of 456 manually annotated Sentinel-2 image tiles acquired between 2017 and 2023 was compiled across diverse emission sources and environmental conditions, with plume boundaries extracted using an MBMP-guided procedure. SimCLR was then used to learn transferable spatial–spectral representations from unlabeled data through contrastive learning, followed by supervised fine-tuning using a unified encoder–decoder segmentation design.

Results across five labeled-data portions (20%, 40%, 60%, 80%, and 100%) confirm that SimCLR pretraining consistently improves segmentation performance relative to ImageNet-only initialization. The strongest performances were achieved by MobileNet, SwinT, and Xception, with MobileNet reaching a recall of 0.92, an IoU of 0.80, and an F1-score of 0.90 at full data

availability, and SwinT achieving a recall of 0.92, an IoU of 0.75, and an F1-score of 0.85. The benefit of SimCLR became more pronounced under limited labels, where ImageNet-only models showed degraded IoU and F1-scores, while SimCLR-pretrained encoders maintained better generalization and produced more physically consistent plume structures in qualitative comparisons.
The study also demonstrated that segmentation quality has a direct influence on CH4 emission quantification using the IME method. Using predicted plume masks, MobileNet produced the most accurate emission estimates. Moreover, uncertainty bounds increased with Qs across all models, consistent with the combined effects of plume boundary variability, retrieval noise, and wind speed uncertainty.

References

Change, I. C. (2013). The physical science basis. (No Title). https://cir.nii.ac.jp/crid/1371413280484207233

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. International Conference on Machine Learning, 1597–1607. http://proceedings.mlr.press/v119/chen20j.html

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1251–1258. http://openaccess.thecvf.com/content_cvpr_2017/html/Chollet_Xception_Deep_Learning_CVPR_2017_paper.html

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (arXiv:2010.11929). arXiv. http://arxiv.org/abs/2010.11929

Etminan, M., Myhre, G., Highwood, E. J., & Shine, K. P. (2016). Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. Geophysical Research Letters, 43(24). https://doi.org/10.1002/2016GL071930

Gorroño, J., Varon, D. J., Irakulis-Loitxate, I., & Guanter, L. (2023). Understanding the potential of Sentinel-2 for monitoring methane point emissions. Atmospheric Measurement Techniques, 16(1), 89–107.

Guanter, L. (2021). Mapping methane point emissions with the PRISMA spaceborne imaging spectrometer. Remote Sensing of Environment, 14.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (arXiv:1704.04861). arXiv. https://doi.org/10.48550/arXiv.1704.04861

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4700–4708. http://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.html

Irakulis-Loitxate, I., Gorroño, J., Zavala-Araiza, D., & Guanter, L. (n.d.). Satellites detect a methane ultra-emission event from an offshore platform in the Gulf of Mexico. 23.

Irakulis-Loitxate, I., Guanter, L., Liu, Y.-N., Varon, D. J., Maasakkers, J. D., Zhang, Y., Chulakadabba, A., Wofsy, S. C., Thorpe, A. K., Duren, R. M., Frankenberg, C., Lyon, D. R., Hmiel, B., Cusworth, D. H., Zhang, Y., Segl, K., Gorroño, J., Sánchez-García, E., Sulprizio, M. P., … Jacob, D. J. (2021). Satellite-based survey of extreme methane emissions in the Permian basin. Science Advances, 7(27), eabf4507. https://doi.org/10.1126/sciadv.abf4507

Irakulis-Loitxate, I., Guanter, L., Maasakkers, J. D., Zavala-Araiza, D., & Aben, I. (2022). Satellites Detect Abatable Super-Emissions in One of the World’s Largest Methane Hotspot Regions. Environmental Science & Technology, 56(4), 2143–2152. https://doi.org/10.1021/acs.est.1c04873

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision, 10012–10022.

https://openaccess.thecvf.com/content/ICCV2021/html/Liu_Swin_Transformer_Hierarchical_Vision_Transformer_Using_Shifted_Windows_ICCV_2021_paper

Marjani, M., Mahdianpari, M., Varon, D. J., & Mohammadimanesh, F. (2025). The integration of vision transformers and SAM for automated methane super-emitter detection using TROPOMI data. Journal of Environmental Management, 393, 127034.

Marjani, M., Mohammadimanesh, F., Varon, D. J., Radman, A., & Mahdianpari, M. (2024). PRISMethaNet: A novel deep learning model for landfill methane detection using PRISMA satellite data. ISPRS Journal of Photogrammetry and Remote Sensing, 218, 802–818.

Mohammadimanesh, F., Mahdianpari, M., Radman, A., Varon, D., Hemati, M., & Marjani, M. (2025). Advancements in satellite-based methane point source monitoring: A systematic review. ISPRS Journal of Photogrammetry and Remote Sensing, 224, 94–112. https://doi.org/10.1016/j.isprsjprs.2025.03.020

Radman, A., Mahdianpari, M., Varon, D. J., & Mohammadimanesh, F. (2023). S2MetNet: A novel dataset and deep learning benchmark for methane point source quantification using Sentinel-2 satellite imagery. Remote Sensing of Environment, 295, 113708. https://doi.org/10.1016/j.rse.2023.113708

Rouet-Leduc, B., Kerdreux, T., Tuel, A., & Hulbert, C. (2023). Autonomous Detection of Methane Emissions in Multispectral Satellite Data Using Deep Learning (arXiv:2308.11003). arXiv. https://doi.org/10.48550/arXiv.2308.11003

Sánchez-García, E., Gorroño, J., Irakulis-Loitxate, I., Varon, D. J., & Guanter, L. (2021). Mapping methane plumes at very high spatial resolution with the WorldView-3 satellite [Preprint]. Gases/Remote Sensing/Data Processing and Information Retrieval. https://doi.org/10.5194/amt-2021-238

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2818–2826. https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.html

Varon, D. J., Jacob, D. J., Jervis, D., & McKeever, J. (2020). Quantifying Time-Averaged Methane Emissions from Individual Coal Mine Vents with GHGSat-D Satellite Observations. Environmental Science & Technology, 54(16), 10246–10253. https://doi.org/10.1021/acs.est.0c01213

Varon, D. J., Jervis, D., McKeever, J., Spence, I., Gains, D., & Jacob, D. J. (2021). High-frequency monitoring of anomalous methane point sources with multispectral Sentinel-2 satellite observations. Atmospheric Measurement Techniques, 14(4), 2771–2785. https://doi.org/10.5194/amt-14-2771-2021

Vaughan, A., Mateo-García, G., Gómez-Chova, L., Růžička, V., Guanter, L., & Irakulis-Loitxate, I. (2023). CH4Net: A deep learning model for monitoring methane super-emitters with Sentinel-2 imagery [Preprint]. Gases/Remote Sensing/Data Processing and Information Retrieval. https://doi.org/10.5194/egusphere-2023-563

Wang, Y., Albrecht, C. M., Braham, N. A. A., Mou, L., & Zhu, X. X. (2022). Self-supervised learning in remote sensing: A review. IEEE Geoscience and Remote Sensing Magazine, 10(4), 213–247.

Zhao, S., Zhang, Y., Zhao, S., Wang, X., & Varon, D. J. (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, 25(7), 4035–4052.