Spectral Channel Attention Network: A Method for Hyperspectral Semantic Segmentation of Cloud and Shadows

Manuel Pérez-Carrasco[*,1] Maya Nasr[2,3] Sébastien Roche[2,3] Chris Chan Miller[2,3]
Zhan Zhang[2,3] Core Francisco Park[4] Eleanor Walker[3]
Cecilia Garraffo[1] Douglas Finkbeiner[4] Ritesh Gautam[2] Steven Wofsy[3]

manuel.perez_carrasco@cfa.harvard.edu*, {mayanasr, sroche, cmiller, ewalker}@g.harvard.edu, {cgarraffo, dfinkbeiner}@cfa.harvard.edu, rgautam@edf.org, wofsy@g.harvard.edu

[1]AstroAI, Center for Astrophysics | Harvard & Smithsonian
[2]Environmental Defense Fund
[3]Department of Earth and Planetary Sciences, Harvard University
[4]Department of Physics, Harvard University

Abstract

Accurate detection of clouds and cloud shadows is essential for reliable atmospheric methane retrievals, a critical component of global climate monitoring efforts. This work presents the Spectral Channel Attention Network (SCAN), a simple deep learning architecture that addresses the fundamental challenge of spectral band selection for hyperspectral cloud and shadow detection through channel-wise attention mechanisms. Unlike traditional approaches that treat all spectral bands equally, SCAN dynamically weights spectral channels based on their discriminative power for atmospheric artifact detection. We evaluate SCAN on MethaneSAT and MethaneAIR hyperspectral datasets, demonstrating superior performance on MethaneSAT (71.53% F1-score vs U-Net's 68.56%). Furthermore, we show that SCAN's spectral attention capabilities can be effectively combined with spatial processing through ensemble approaches, achieving the best results with F1-scores of 78.50% for MethaneAIR and 78.80% for MethaneSAT.

Our code is publicly available at: https://doi.org/10.7910/DVN/IKLZOJ

1 Introduction

Remote sensing of atmospheric greenhouse gases has emerged as a critical tool for climate monitoring and emission quantification. Methane (CH_4) presents a particularly urgent target for climate mitigation efforts due to its high warming potential—over 80 times that of CO_2 during the first two decades after emission [1, 2]. This makes accurate methane monitoring essential for global climate goals.

The MethaneSAT mission [3] and its airborne companion MethaneAIR [4, 5, 6, 7, 8] represent a new generation of hyperspectral imaging spectrometers designed for precise methane quantification. These platforms enable detailed quantification of both point sources and area emissions at unprecedented spectral and spatial resolution. However, accurate retrieval of methane concentrations faces a challenge: clouds and cloud shadows introduce significant artifacts that bias atmospheric retrievals.

Traditional cloud detection methods typically treat all spectral bands equally, which may not fully exploit the spectral heterogeneity inherent in hyperspectral data [9, 10, 11, 12, 13, 14, 15, 16]. While

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2025.

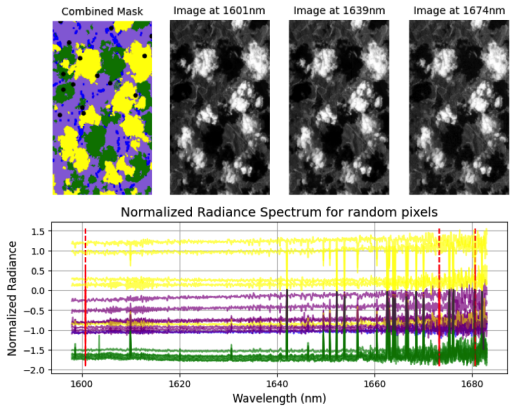


Figure 1: MethaneAIR data example with classification mask (purple: background, yellow: clouds, green: shadows, blue: dark surfaces) and three spectral wavelength images (top). Bottom panel shows normalized radiance spectra from 10 randomly sampled soundings per class.

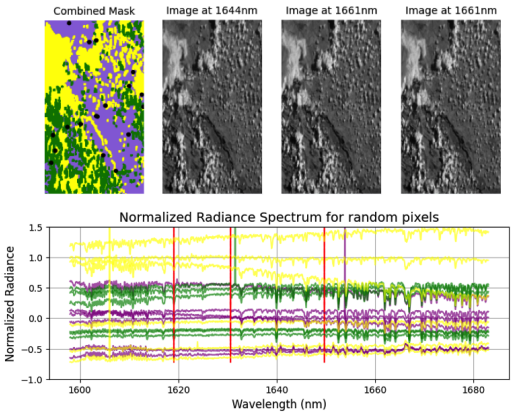


Figure 2: MethaneSAT data example with classification mask (purple: background, yellow: clouds, green: shadows) and three spectral band images (top). Bottom panel shows normalized radiance spectra from 10 randomly sampled soundings per class.

deep learning approaches like U-Net have demonstrated strong performance for spatial segmentation [17, 18, 19, 20, 21, 22, 23, 24], they face challenges when applied to hyperspectral data where different spectral regions provide varying levels of discrimination between atmospheric artifacts and surface materials.

This paper presents the Spectral Channel Attention Network (SCAN), a simple deep learning architecture that addresses spectral band selection through channel-wise attention mechanisms. SCAN dynamically weights spectral bands based on their discriminative features for cloud and shadow detection. We demonstrate that SCAN's spectral attention capabilities can be effectively combined with U-Net's spatial strengths through simple ensemble approaches.

We evaluate SCAN against baseline methods on MethaneSAT and MethaneAIR datasets, showing superior performance on MethaneSAT, and competitive results on MethaneAIR. Furthermore, our simple ensemble method achieves F1-scores of 78.50±3.08% for MethaneAIR and 78.80±1.28% for MethaneSAT. Our contributions are threefold: (1) we show that applying channel attention directly to spectral bands—rather than to spatial feature maps—provides significant benefits for hyperspectral atmospheric artifact detection; (2) we provide comprehensive evaluation on real MethaneSAT and MethaneAIR data, showing that domain-informed spectral weighting outperforms general-purpose mechanisms; and (3) we show that simple ensemble strategies combining spectral and spatial processing achieve the best results for both MethaneAIR and MethaneSAT by margins of 4% and 7% of F1-score respectively.

2 Data and Preprocessing

Our study utilizes calibrated and georeferenced Level 1B (L1B) hyperspectral data from MethaneSAT and MethaneAIR O_2 spectrometers. Figure 1 and Figure 2 shows some sample images for both instruments. The MethaneAIR [5, 6, 7, 8] dataset comprises 508 hyperspectral cubes with 1024 spectral bins covering 1592-1678 nm, spatially cropped to $\sim$300$\times$178 spatial soundings. MethaneSAT data contains 262 samples with 1080 spectral bins covering 1598-1683 nm at $\sim$220$\times$200 km^2 coverage. Ground truth masks for each L1B sample are derived using Level2 (L2) retrieved quantities, including CO_2 and CH_4 vertical column densities (VCDs) from the CH_4 spectrometer and surface pressure (P) from the O_2 spectrometer (see [5] for details). Masks contain categories clouds, cloud shadows, dark surfaces, and background for MethaneAIR, and clouds, clouds shadows, and background categories for MethaneSAT

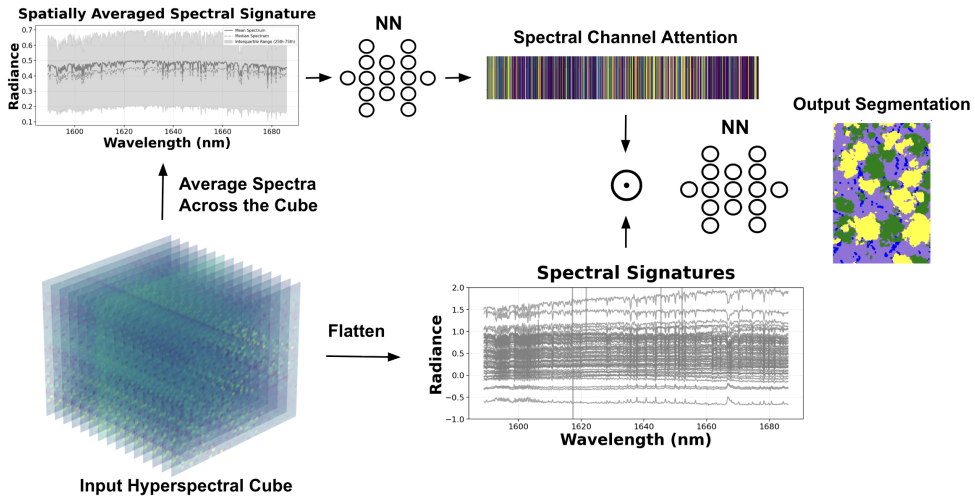


Figure 3: Overview of the Spectral Channel Attention Network (SCAN) architecture for hyperspectral cloud and shadow detection. The input hyperspectral cube is flattened and processed through a neural network to extract spectral signatures for each pixel. These signatures are then weighted by the spectral channel attention module, which learns to emphasize discriminative wavelengths while suppressing less informative spectral regions. The attended features are subsequently fed into a classification network to produce the final segmentation map distinguishing background (purple), clouds (yellow), and cloud shadows (green).

Preprocessing includes missing value imputation, spatial standardization, and two-step normalization: percentile clipping (1st-99th) followed by per-band standardization and batch-wise normalization across all dimensions.

3 Spectral Channel Attention Network

The key insight driving our approach is that effective hyperspectral cloud and shadow detection requires dynamic spectral band selection, emphasizing wavelengths that are most informative while suppressing less discriminative regions.

Inspired from [25], we propose Spectral Channel Attention Network (SCAN), a simple neural network that adapts channel-wise attention mechanisms for hyperspectral band selection in cloud and shadow detection. An overview of our proposed method is shown in 3

Spectral Attention Module: Given input hyperspectral data $X \in \mathbb{R}^{B\times H\times W\times C}$, where B is batch size, H and W are spatial dimensions, and C is the number of spectral channels, we compute channel-wise attention weights as:

$$\alpha = \sigma(W_2 \text{ReLU}(W_1 \bar{x})) \tag{1}$$

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

$$X_{att} = X \odot \alpha \tag{2}$$

where $\odot$ broadcasts attention weights across all spatial locations, producing a spectrally-weighted representation that emphasizes discriminative bands while suppressing less informative spectral regions.

Classification Framework: The attended features are processed through a fully-connected network for pixel-wise classification:

$$P(y|X) = \text{Softmax}(f_{MLP}(X_{att})) \tag{3}$$

Ensemble Methods To leverage complementary strengths of spectral and spatial processing, we develop two simple ensemble methods that combine SCAN and U-Net strenghts by concatenating predictions from frozen pre-trained models. We processes predictions through an Multilayer Perceptron (MLP; [26]) with hidden layers [256, 128] and dropout ($\delta = 0.2$). Also, we use a Convolutional Neural Network (CNN; [27]) composed of 3×3 convolutional layers with padding 1 and channel dimensions [64, 128, 256], followed by 1×1 convolution for classification to preserve spatial relationships

Relationship to Channel Attention Mechanisms: SCAN's spectral attention module shares conceptual similarities with well established channel attention mechanisms, particularly Squeeze-and-Excitation Networks (SE-Net) [28], Efficient Channel Attention (ECA-Net) [29], and Convolutional Block Attention Module (CBAM) [30]. Similar to these methods, our approach employs global average pooling followed by a bottleneck neural network to generate channel-wise weights. However, we show that when applied to the spectral dimension of hyperspectral data for atmospheric artifact detection, significant improvements are obtained over both spatial-spectral approaches and more complex self-attention mechanisms. While similar approaches are designed for natural image classification with spatial feature maps, we adapt this principle to weight spectral bands in hyperspectral cubes, where channels represent physical wavelengths rather than learned feature maps.

4 Results

Baselines: We evaluate SCAN against three baseline approaches across MethaneSAT and MethaneAIR datasets. Iterative Logistic Regression (ILR; [10]), MLP [31]), and U-Net [18], as well with our ensemble methods (Combined MLP and Combined CNN).

Training and Evaluation Strategy: All models use weighted cross-entropy loss with inverse class frequency weights and Adam optimization with dataset-specific learning rates determined via 3-fold cross-validation (See Appendix A for a details). Training includes data augmentation (random flips and rotations), 100 epochs with batch size 32, and early stopping after 20 epochs without validation improvement. For MethaneSAT's variable dimensions, we implement patch-based evaluation using overlapping 224×224 patches with weighted averaging. We report our final results over a separate test set (composed by the 20% of the data) and the best performing model checkpoints based on validation performance of each fold.

Quantitative Results: Table 1 presents comprehensive performance metrics. SCAN demonstrates strong spectral attention capabilities, achieving notable success particularly on MethaneSAT data where it outperforms U-Net with 80.33±3.43% accuracy and 71.53±0.75% F1-score compared to U-Net's 78.73±3.23% accuracy and 68.56±0.36% F1-score. This superior performance on MethaneSAT highlights SCAN's effectiveness in leveraging spectral band selection for datasets where spectral discrimination is critical.

	MethaneAIR		MethaneSAT	
Model	Acc (%)	F1 (%)	Acc (%)	F1 (%)
Individual Methods				
ILR	73.81±4.05	62.07±0.86	71.82±4.02	64.35±3.56
MLP	82.49±2.24	71.29±1.02	74.03±3.72	67.11±2.06
U-Net	**88.26±0.45**	**76.24±1.90**	78.73±3.23	68.56±0.36
SCAN	86.51±2.90	74.96±0.96	**80.33±3.43**	**71.53±0.75**
Ensemble Methods				
Combined MLP	88.92±1.80	76.99±6.78	81.32±1.28	78.10±1.72
Combined CNN	**89.42±1.20**	**78.50±3.08**	**81.96±1.45**	**78.80±1.28**

Table 1: Performance comparison across different models for both datasets.

For MethaneAIR data, while SCAN achieves 86.51±2.90% accuracy and 74.96±0.96% F1-score—slightly below U-Net's performance—the difference is modest (1.75% accuracy, 1.28% F1-score). This demonstrates that SCAN's spectral attention approach remains competitive with

spatial methods even when spatial features may be more discriminative. The Combined CNN method achieves state-of-the-art results on both datasets: 89.42±1.20% accuracy and 78.50±3.08% F1-score for MethaneAIR, and 81.96±1.45% accuracy and 78.80±1.28% F1-score for MethaneSAT. Detailed model comparisons can be found in Appendix B.

Comparison with Attention Mechanisms: To further validate the suitability of our algorithm, we compared SCAN against both channel attention and transformer-based self-attention methods on MethaneSAT data. For channel attention, we implemented SE-UNet, augmenting U-Net with Squeeze-and-Excitation blocks [28] applied to spatial feature maps. For self-attention, we evaluated Vision Transformer [32] with SegFormer head [33] (applying standard self-attention to spectral-spatial patches) and a spectral transformer approach based on SpectralFormer [34]. Specifically, we created patch embeddings by grouping neighboring spectral bands (30 bands per patch), projecting them to fixed embedding size, and using a transformer with class token for pixel classification.

SCAN substantially outperforms all attention-based approaches, achieving 71.53% F1-score compared to SE-UNet's 61.10% F1-score, ViT-SegFormer's 60.97% F1-score, and SpectralFormer's 62.65% F1-score (See Appendix C for details). These results show large performance gaps of improvement over SE-UNet, Spectralformer and ViT-SegFormer, validating that SCAN's spectral attention mechanisms operating directly on physical wavelengths is suitable for hyperspectral cloud and shadow segmentation tasks.

5 Conclusion

This work presents the Spectral Channel Attention Network (SCAN), a novel architecture for hyperspectral cloud and shadow detection that addresses the fundamental challenge of spectral band selection through channel-wise attention mechanisms. SCAN demonstrates superior performance on MethaneSAT data, outperforming U-Net with 71.53% F1-score versus 68.56%, while remaining competitive on MethaneAIR data. Notably, SCAN substantially outperforms transformer-based self-attention methods, achieving 11.12% and 13.99% F1-score improvements over spectral and spatial self-attention approaches respectively.

The effectiveness of SCAN can be further enhanced through simple ensemble strategies that combine spectral attention with spatial processing. Our Combined CNN approach achieves state-of-the-art performance with F1-scores of 78.50% for MethaneAIR and 78.80% for MethaneSAT, representing significant improvements over baseline methods. These advances in atmospheric artifact detection directly enhance satellite-based methane monitoring reliability, supporting critical climate monitoring efforts. SCAN's focused approach to spectral band weighting proves that specialized attention mechanisms are more effective than general methods that process bands equally for hyperspectral segmentation tasks, opening new directions for hyperspectral remote sensing applications.

6 Acknowledgments

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[27] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[28] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

[29] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11534–11542, 2020.

[30] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[31] Anita Thakur and Deepak Mishra. Hyper spectral image classification using multilayer perceptron neural network & functional link ann. In *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, pages 639–642, 2017.

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

[33] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090, 2021.

[34] Danfeng Hong, Zhu Han, Jing Yao, Lianru Gao, Bing Zhang, Antonio Plaza, and Jocelyn Chanussot. Spectralformer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2021.

A Appendix A: Hyperparameter Selection

We performed hyperparameter optimization using 3-fold cross-validation on 10% of the training data. Learning rates were selected from $\{1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 5 \times 10^{-3}, 1 \times 10^{-2}\}$, using 3-fold cross-validation on the training set. The optimal rates are shown in Table 2.

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

Table 2: Best learning rates for each data source and model architecture.

All models used Adam optimizer $\beta_1 = 0.9$ and $\beta_2 = 0.999$ with class weighting based on inverse class frequencies. Key architectural parameters: MLP hidden dimensions [20,20], SCAN with channel reduction of 16 and MLP classifier of dimensions [20, 20] , Combined MLP dimensions [256,128] with 0.2 dropout, Combined CNN channels [64,32,16] with 0.2 dropout.

B Appendix B: Detailed Baseline Results

The visual comparison shown in Figure 4 and 5 reveals distinct model characteristics across both datasets. ILR and MLP models produce noisy, fragmented predictions due to pixel-wise processing, while SCAN demonstrates improved boundary detection through spectral attention mechanisms. U-Net reduced dark surface overprediction, though its lower quantitative metrics reflect its tendency to boundary over-smoothing. The Combined CNN achieves optimal balance, preserving structural detail while maintaining spatial coherence across both MethaneAIR and MethaneSAT scenes.

Confusion matrices in Figure 6 for MethaneAIR and Figure 7 for MethaneSAT reveal dataset-specific classification challenges. MethaneAIR shows progression from ILR's background-dark surface confusion (18.09% misclassification) to Combined CNN's balanced performance across all classes (91.50% shadow accuracy). MethaneSAT presents greater cloud-shadow spectral similarity, with SCAN outperforming U-Net in shadow detection (81.62% vs 71.11%), highlighting spectral attention benefits for this dataset's characteristics.

For MethaneAIR data, the U-Net emerged as the second-best performing model due to its capacity for detecting dark surfaces and producing spatially coherent predictions with reduced false positives. The model achieved 88.26±0.45% accuracy while demonstrating superior background detection (86.71%) and notably reducing dark surface misclassifications. In contrast, SCAN showed strong cloud detection capabilities (91.20% class-specific accuracy) but struggled with shadow-dark surface discrimination. For MethaneSAT data, SCAN outperformed U-Net (accuracy: 80.33±3.43% vs 78.73±3.23%), suggesting that spectral attention mechanisms are particularly valuable for MethaneSAT's unique spectral characteristics, especially in shadow detection where SCAN achieved 81.62% accuracy compared to U-Net's 71.11%.

The confusion matrices revealed distinct classification challenges between datasets. MethaneAIR models showed primary confusion between background and dark surface classes, with the Combined CNN reducing this to acceptable levels while achieving excellent cloud (96.26%) and shadow (91.50%) detection. MethaneSAT presented greater spectral similarity between cloud and shadow classes, with even the best-performing Combined CNN showing 12.59% cloud-to-shadow misclassification. However, the Combined CNN maintained robust overall performance with class-specific accuracies of 85.02% for background, 79.83% for clouds, and 83.15% for shadows. The optimal

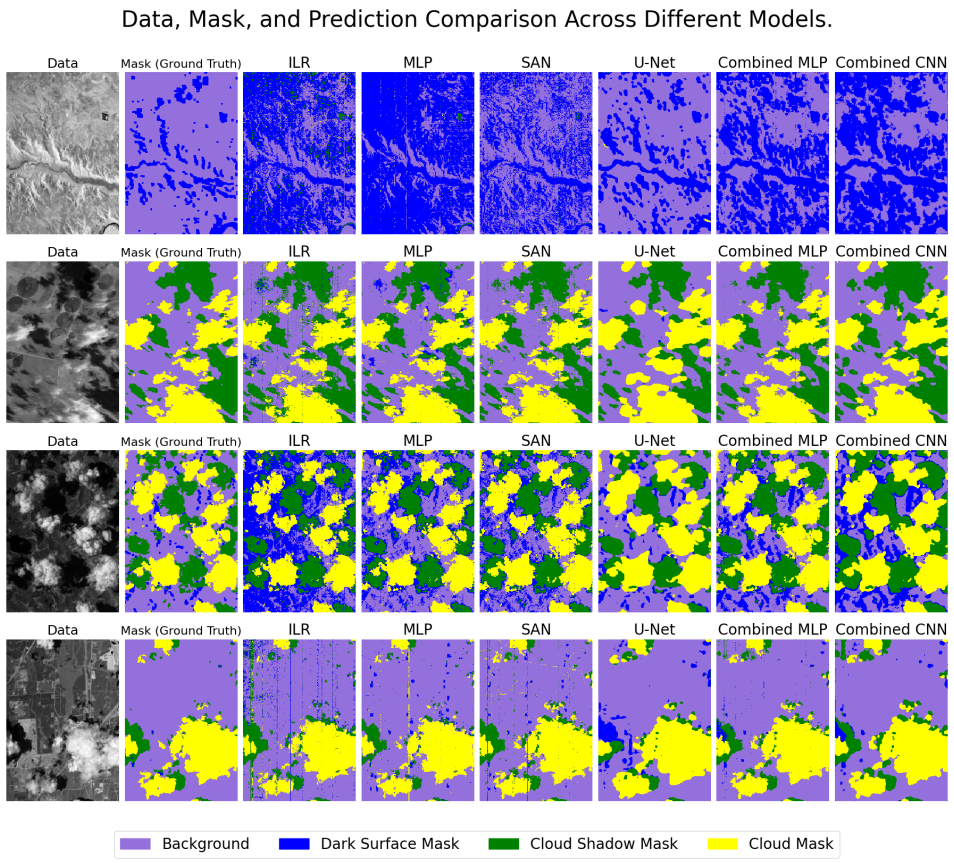


Figure 4: Prediction comparison across all evaluated models for MethaneAIR test scenes from first cross-validation fold.

model selection should be guided by specific operational requirements, as U-Net's tendency toward conservative boundary delineation may be advantageous for applications prioritizing false positive reduction over precise edge detection, particularly for downstream atmospheric retrieval processes.

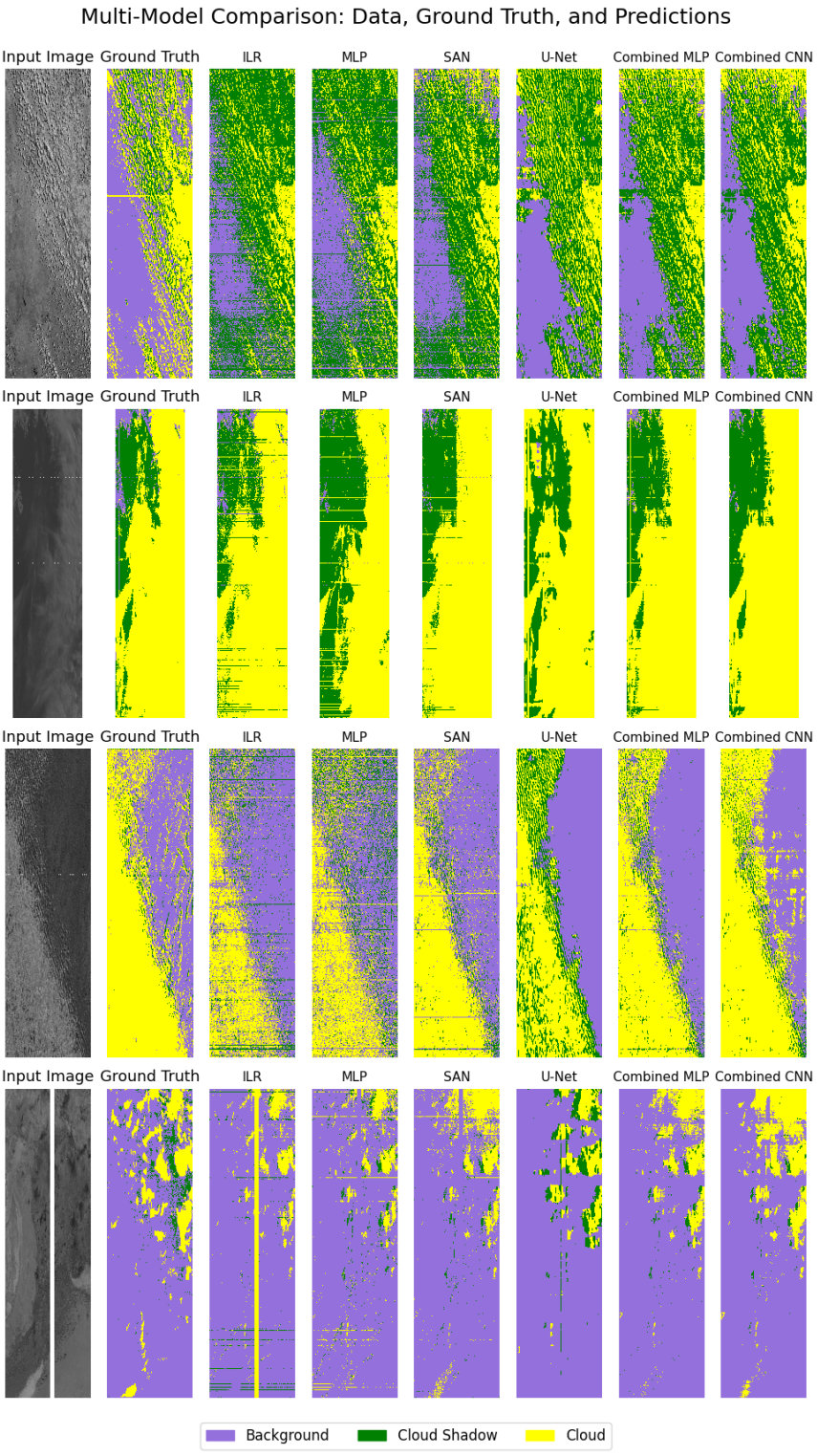
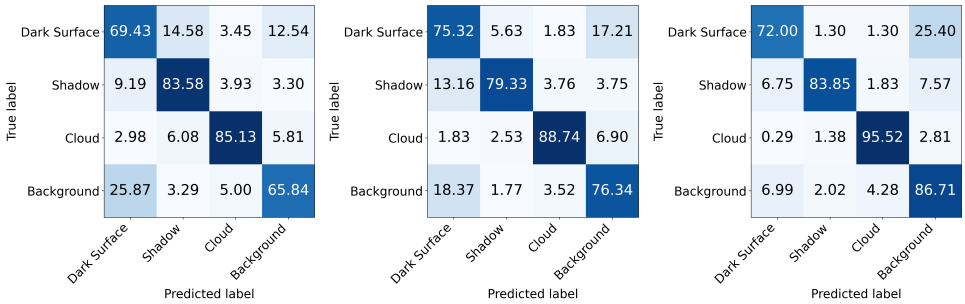
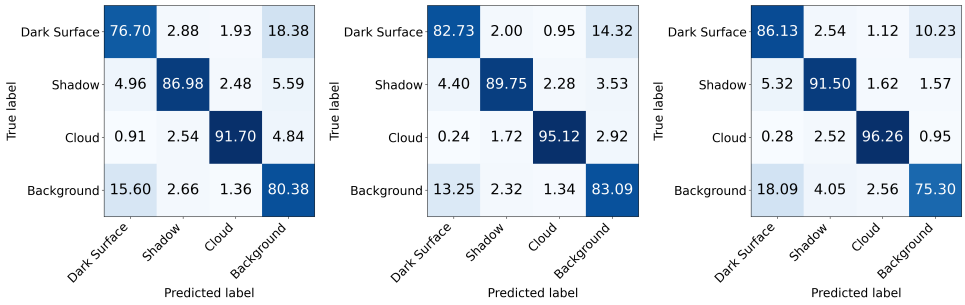


Figure 5: Prediction comparison across all evaluated models for MethaneSAT test scenes from first cross-validation fold. Scenes transposed for visualization.

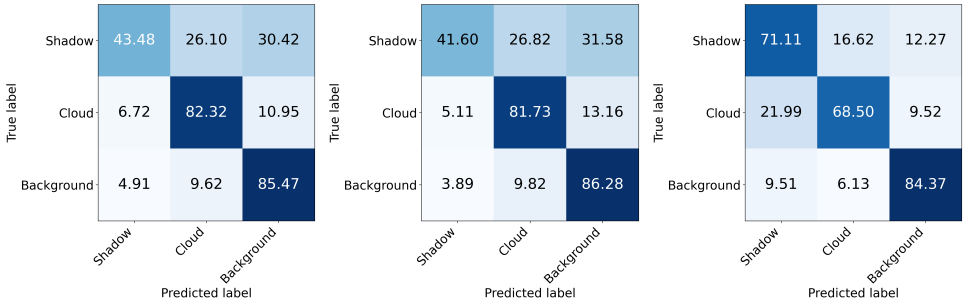


(a) ILR (b) MLP (c) U-Net

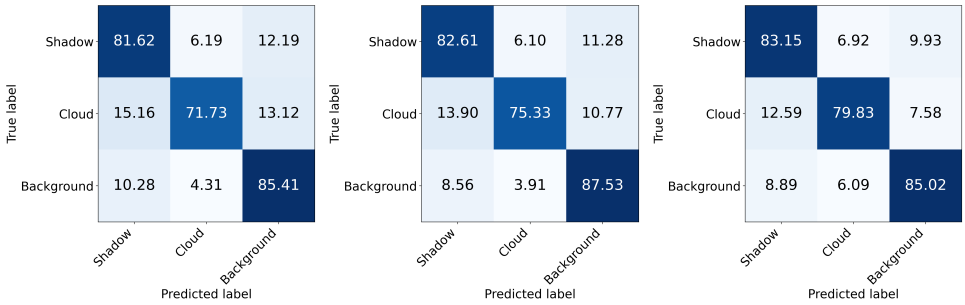


(d) SCAN (e) Combined MLP (f) Combined CNN

Figure 6: Confusion matrices for all evaluated models on MethaneAIR test data from first cross-validation fold.



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



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

Figure 7: Confusion matrices for all evaluated models on MethaneSAT test data from first cross-validation fold.

C Appendix C: Attention Mechanism Results

To evaluate the effectiveness of spectral channel attention against alternative attention mechanisms, we conducted experiments with three baseline approaches on MethaneSAT data: SE-UNet (spatial

channel attention), ViT-SegFormer (spatial self-attention), and SpectralFormer (spectral self-attention). This appendix provides detailed quantitative and qualitative results for these comparisons.

Table 3 presents the quantitative performance of all attention-based methods. SE-UNet applies Squeeze-and-Excitation blocks to U-Net's spatial feature maps, achieving 72.81±8.24% accuracy but exhibiting high variance (±14.2% F1-score). The transformer-based approaches show lower but more stable performance, with ViT-SegFormer achieving 68.84±2.15% accuracy and SpectralFormer achieving 70.41±2.94% accuracy. SCAN substantially outperforms all methods with 80.33±3.43% accuracy and 71.53±0.75% F1-score, while maintaining the lowest variance across all metrics.

Model	**Acc (%)**	**F1 (%)**
SE-UNet	72.81±8.24	61.10±14.2
ViT-SegFormer	68.84±2.15	60.97±3.42
SpectralFormer	70.41±2.94	62.65±1.69
SCAN	**80.33±3.43**	**71.53±0.75**

Table 3: Performance comparison of attention mechanisms on MethaneSAT test data.

Figure 8 shows representative MethaneSAT test scenes comparing attention-based baseline methods. SE-UNet produces the most spatially coherent predictions among other attention methods, with well-defined cloud and shadow boundaries, but exhibits visible cloud-shadow confusion (visible in rows 1 and 3). ViT-SegFormer exhibits severe limitations, producing highly fragmented and noisy predictions across all test scenes. Notable failures include nearly complete misclassification in row 1 (predicting mostly shadows where clouds dominate), and general difficulties to capture spatial structures. SpectralFormer, produces noisy predictions with limited spatial coherence, failing to capture cloud and shadows. The pervasive noise indicates that spectral self-attention alone, without spatial regularization, leads to unstable pixel-wise predictions that struggles distinguishing atmospheric features from noise.

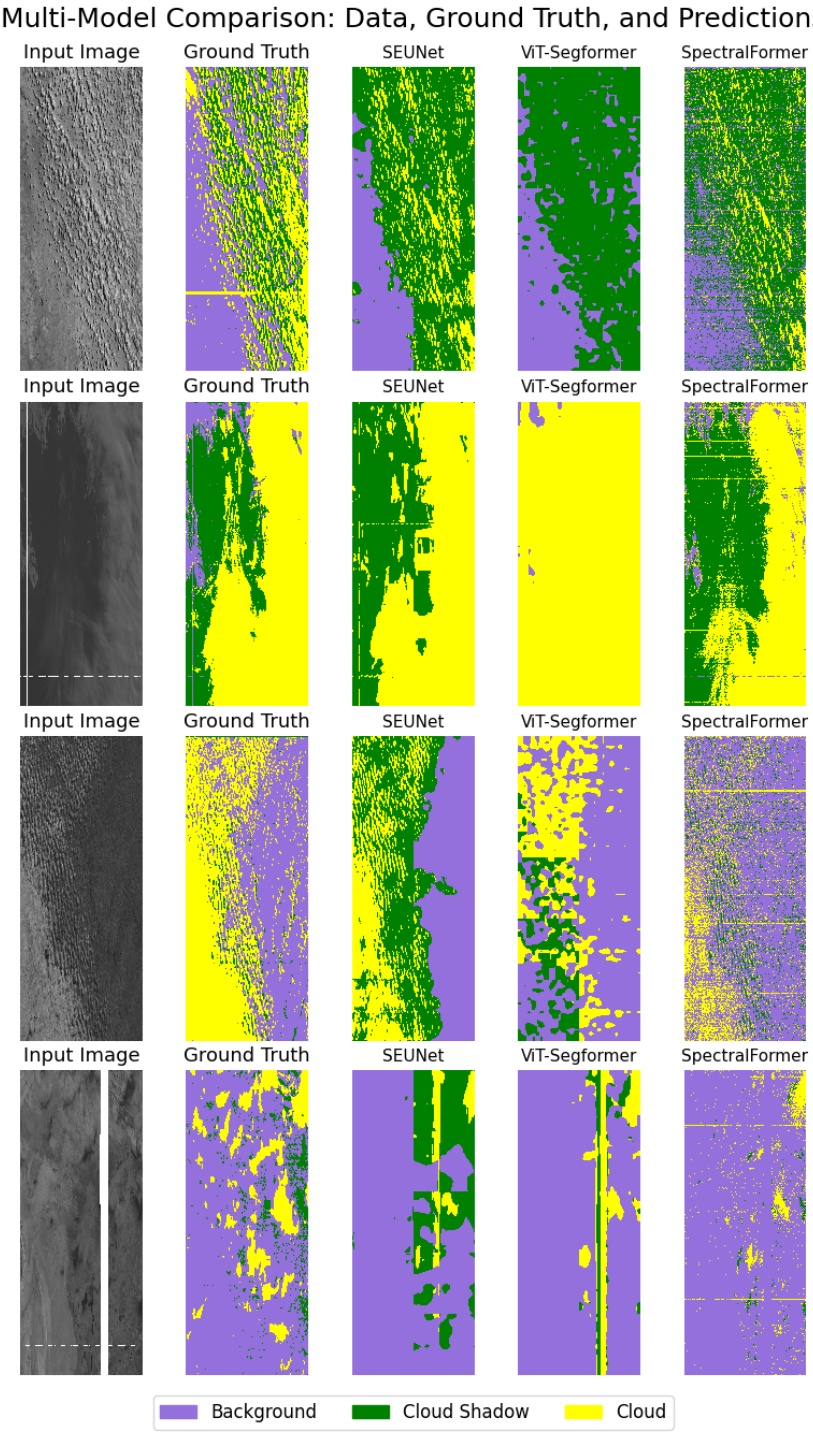


Figure 8: Prediction comparison of attention-based methods on MethaneSAT test scenes. From left to right: Input image, Ground Truth, SE-UNet, ViT-SegFormer, and SpectralFormer.