HyperFM: An Efficient Hyperspectral Foundation Model with Spectral Grouping

Supplementary Material

8. Our HyperFM250k Dataset

Hyperspectral imaging from space offers detailed spectral information about the Earth's surface and atmosphere, and recent missions have significantly increased the volume and quality of available data. Notable examples include EnMAP [17], PRISMA [35], and the forthcoming CHIME mission [31]. These systems are optimized for land-focused applications and provide the spatial, spectral, and radiometric performance required for tasks such as vegetation characterization, mineralogical mapping, and terrestrial environmental monitoring. Their design, however, limits their suitability for marine, cryospheric, and atmospheric observations. Over ocean and polar regions, the signal is weaker and more sensitive to calibration uncertainties, while atmospheric studies often require broader spectral coverage, higher radiometric stability, and measurement strategies tailored to cloud and aerosol retrievals. Consequently, although current missions substantially advance land-oriented hyperspectral remote sensing, they provide only partial support for applications central to climate research and numerical weather prediction, underscoring the need for instruments explicitly designed for these domains.

NASA's PACE mission directly addresses the limitations of land-oriented hyperspectral systems by introducing the Ocean Color Instrument (OCI), which delivers global hyperspectral coverage tailored to oceanic and atmospheric applications. OCI provides continuous spectral sampling from roughly 340 to 890 nm at 5 nm resolution, along with seven additional shortwave infrared channels centered at 940, 1038, 1250, 1378, 1615, 2130, and 2260 nm. This combination offers the spectral density and calibration stability required for robust retrievals of ocean color, cloud properties, and aerosol composition. In contrast, earlier multispectral instruments such as MODIS [36] and VIIRS [37] have supported long-term atmospheric and ocean observations but lack the spectral detail and radiometric performance needed for reliable microphysical characterization. As a result, PACE OCI functions as a complementary capability to existing hyperspectral missions while providing the spectral depth necessary for improved environmental and climate research.

8.1. Dataset Overview

HyperFM250k is a hyperspectral dataset built from PACE-OCI observations. It contains 250k HSI tiles with cloud coverage more than 60%, addressing the limitations of earlier releases that often kept the cloud presence below ten percent.

8.1.1. Raw Data Overview

A summary of the raw PACE-OCI data used to generate **HyperFM250k** is shown in Tab. 4. PACE-OCI provides Level-1 radiance data in three forms: A) raw, B) calibrated and geolocated, and C) calibrated, geolocated, and co-registered to a common grid. Because the raw measurements require additional calibration and are not reliable for direct analysis, Level-1B and Level-1C are the practical options. However, level 1C has a coarse spatial resolution of 5km while level 1B offers finer resolution of 1.2km. Therefore we select level 1B observations to construct our dataset. An additional advantage of Level-1B is its colocation with Level-2 products which simplifies downstream use without requiring further preprocessing and alignment. Note that all geophysical science products are released as level-2 products. Some of the key geophysical products include Cloud Optical Thickness (COT), Cloud Effective Radius (CER), Cloud Water Path (CWP), and Cloud Top Height (CTH), among others [29].

A total of 2262 granules were downloaded using a python script specifying the search criterion. The granules were selected from May 2024 to April 2025. The *train* split comprises 12 days of data from 12 different months, while the validation and test splits each use 2 days. The level-1B raw granules totaled approximately $3,625$ GB while level-2 granules were approximately 181 GB.

8.1.2. Preprocessed Dataset Overview

Both Level 1B and Level-2 raw granules are preprocessed using the Algorithm 1. Preprocessed hyperspectral tiles were saved in (.npy) format. A summary of the HyperFM250k is as follow:

- 253,104 hyperspectral tiles. Each tile has input and corresponding target data.
- Train / Val / Test: 168,394 /42,355 /42,355
- Train / Val / Test: 1688.7 GB / 431.73 GB / 431.73 GB
- Each Tile: NumPy Array shaped: $291 \times 96 \times 96$
- Per-channel mean and standard deviation provided

8.2. Tasks

The **HyperFM250k** dataset includes four pixel-wise regression tasks and one binary segmentation task.

- Cloud Optical Thickness (COT) – Also called cloud optical depth. It measures how much incoming solar radiation is attenuated as it passes through a cloud, with larger values indicating optically thicker clouds.

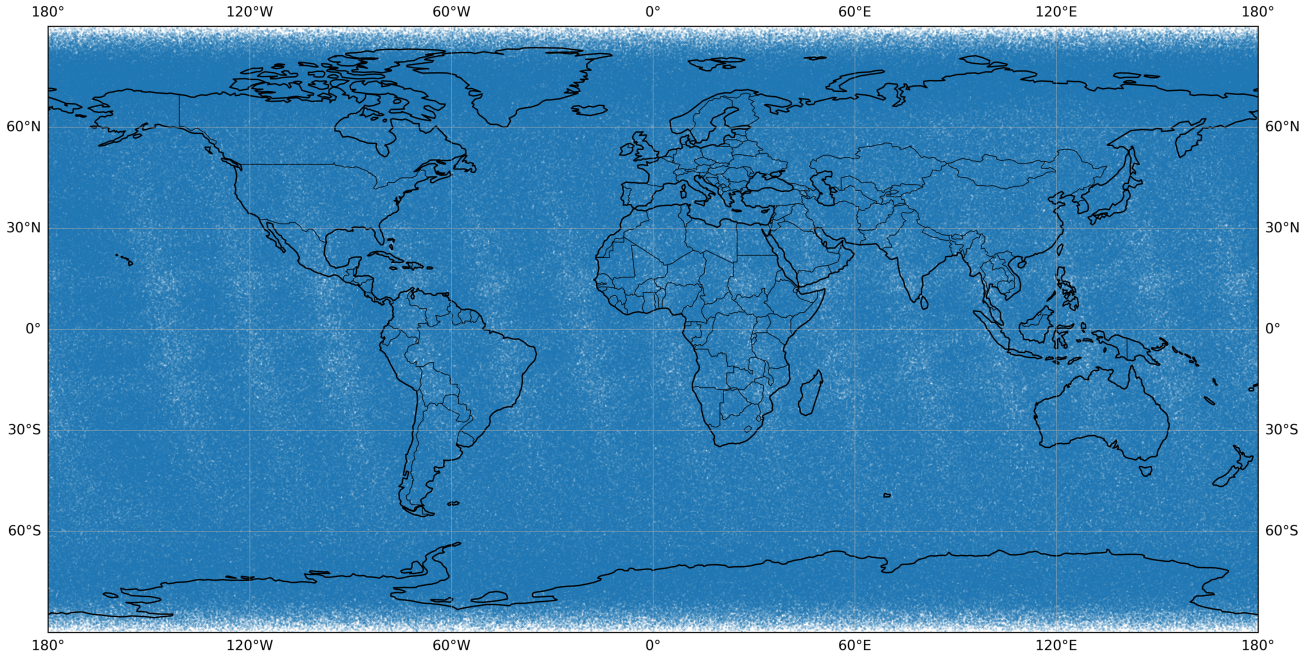

Figure 7. Geographical scatter plot of **HyperFM250k**. The map depicts the global coverage of hyperspectral images within our dataset, demonstrating its extensive geographical scope.

Table 4. Summary of PACE OCI Data used to construct **HyperFM250k** dataset

Product	Processing Level	Variables Used	Spectral Bands	Temporal Coverage	# Raw Granules
PACE_OCI_L1B_SCI	Level-1B	observation_data: rhot_blue, qual_blue rhot_red, qual_red, rhot_SWIR, qual_SWIR	blue (#119): 314∼ 605nm red (#163): 600∼ 895nm SWIR (#9): 940∼ 2258nm Total: #291	Train: 2024-5-10, 2024-6-10, 2024-7-10 2024-8-10, 2024-9-11, 2024-10-10 2024-11-10, 2024-12-10, 2025-1-10 2025-2-10, 2025-3-10, 2025-4-10 Valid: 2025-2-28, 2025-4-28 Test: 2025-2-20, 2025-3-20	Train: 1677, Valid: 291, Test: 294
PACE_OCI_L2_CLOUD	Level-2	geophysical_data: cot_21, cer_21 cwp, cth			Train: 1687, Valid: 291, Test: 294
PACE_OCI_L2_CLOUD_MASK	Level-2	geophysical_data: cloud_flag cloud_flag_dilated			Train: 1687, Valid: 291, Test: 294

- Cloud Effective Radius (CER) – Describes the mean droplet size within the cloud and is closely linked to cloud albedo and precipitation processes.
- Cloud Water Path (CWP) – Represents the vertically integrated liquid water content, providing an estimate of total cloud water and its radiative influence.
- Cloud Top Height (CTH) – Indicates the altitude of the cloud's upper boundary.
- Cloud Mask (binary segmentation) – Identifies whether each pixel is cloud or non-cloud, supporting cloud screening and downstream retrieval quality.

These five tasks capture core aspects of cloud structure and behavior, and they play a central role in climate monitoring [15], weather forecasting [16], aviation safety [18], and renewable energy studies [57].

Due to time and space limitations, we were only able to report on the four regression tasks.

9. Additional Results

We present additional results from Sec. 6 here which were excluded due to space limitation. We compared with another recent hyperspectral foundation model called HyperFree [20] by loading their ViT-base weights and adding the convolutional decoder as shown in Fig. 5. Note that we remove the *neck* from HyperFree ViT-b encoder for fair comparison with all the HFMs. The model was finetuned in the same way the other HFMs [see Sec. 5] were, that is for 250 epochs with an early stopping criterion set to validation MSE loss and a patience of 30. Furthermore, we replaced the Hypoformer block in our HyperFM with regular ViT blocks and called it HyperFM-ViT. We did this to see the trade-off between using *hypoformer* and ViT blocks. HyperFM-ViT are finetuned both decoder only and full network with the same hyperparameters. All the results are reported in Tab. 5. We can see that our *HyperFM* outperforms *HyperFree* along with other variants of HyperFM. The results entail that ViT blocks also achieve similar re-

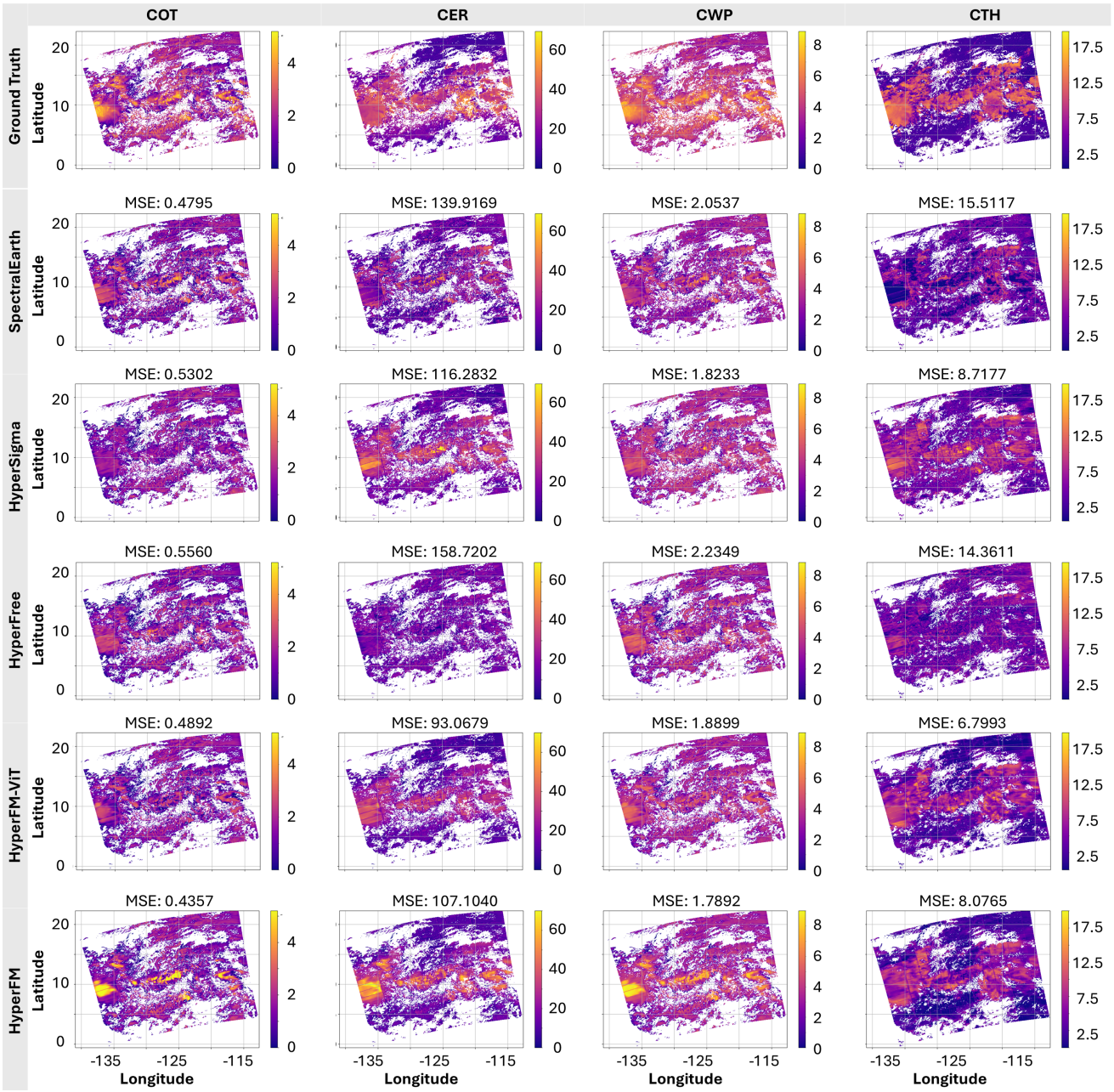

Figure 8. Comparison of Hyperspectral FMs on four pixel-wise regression tasks: cloud optical thickness(COT), cloud effective radius (CER), cloud water path (CWP) and cloud top height (CTH). MSE score for each retrievals are reported on top of the retrieval image. Sometimes grids are noticeable at the patch borders which reduces with overlapped sampling at additional computational cost.

sults but with additional computational cost. It also entails that the key advantage of our model stems from the *Group Embed* module while *Hypoformer* blocks keep it parameter efficient.

Fig. 8 provides a qualitative comparison of all the SOTA HFMs: SpectralEarth [3], HyperSigma [49], HyperFree [20], HyperFM-ViT, and **HyperFM**. For fairness, all retrieval examples are drawn from the decoder-only fine-tuned models. Sec. 6 discusses the SpectralEarth and HyperSigma results with respect to our HyperFM. Beside that we can see that HyperFree shows clear difficulty in recovering the CER property same as *SpectralEarth* and consistently falls behind other hyperspectral foundation models, including SpectralEarth, HyperSigma, and HyperFM in COT, CWP and CTH retrievals. Its design choice to suppress cloudy pixels further limits its ability to handle cloud-centric tasks, which explains the performance gap observed here.

Table 5. Additional Benchmark Results of HyperFM on Four Pixel-wise Regression Tasks

Model Architecture	COT MSE (↓)	CER MSE (↓)	CWP MSE (↓)	CTH MSE (↓)	Model Capacity	Trainable Param	Train Time (Hr)
UNet [19, 30]	0.3928	84.7329	1.5709	7.6786	31.04M	31.04M	1.57@(1x48GB A6000)
CloudUNet [47]	0.3752	128.4650	1.7030	12.1648	0.1.86M	0.1.86M	3.67(@1x12GB RTX2080Ti)
CAM [48]	0.3367	74.4473	1.5123	6.6306	0.47M	0.47M	3.71(@1x12GB RTX2080Ti)
SpectralEarth [3][w/o FT]	61.0280	14384.0316	26.1292	1296.5026	88.78M	-	-
HyperSigma [49][w/o FT]	53.1539	39999.7969	9.1567	979.5114	100.16M	-	-
HyperFree [20][w/o FT]	9.170	21296.765	36.923	1681.148	101.82M	-	-
SpectralEarth [3]	0.4699	97.9197	1.7111	7.6742	88.78M	0.54M	9.58@(3x48GB RTX8000)
HyperSigma [49]	0.3212	95.4874	1.3317	8.4936	100.16M	0.69M	133.31@(2x12GB 2080Ti)
HyperFree [20]	0.5570	117.8959	2.0632	10.0716	101.82M	0.69M	5.38@(2x48GB A6000)
HyperFM-ViT-Dec Only FT	0.3345	74.2573	1.2792	5.4212	49.54M	1.48M	6.07@(3x48GB RTX8000)
HyperFM-ViT-Full FT	0.2672	59.3235	1.0384	4.2518	49.54M	49.54M	3.53@(3x48GB RTX8000)
HyperFM-Dec Only FT [ours]	**0.3124**	**73.7017**	**1.2229**	**5.0976**	32.06M	1.48M	7.50@(2x48GB A6000)
HyperFM-Full FT [ours]	**0.2615**	**62.3963**	**1.0137**	**4.0508**	32.06M	32.06M	5.02@(2x48GB A6000)

References

[1] Peri Akiva, Matthew Purri, and Matthew Leotta. Self-supervised material and texture representation learning for remote sensing tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8203–8215, 2022. 2

[2] Muhammad Awais, Muzammal Naseer, Salman Khan, Rao Muhammad Anwer, Hisham Cholakkal, Mubarak Shah, Ming-Hsuan Yang, and Fahad Shahbaz Khan. Foundation models defining a new era in vision: a survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 2

[3] Nassim Ait Ali Braham, Conrad M Albrecht, Julien Mairal, Jocelyn Chanussot, Yi Wang, and Xiao Xiang Zhu. Spectralearth: Training hyperspectral foundation models at scale. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025. 1, 2, 3, 5, 6, 8, 4

[4] NASA Goddard Space Flight Center. Pace ocean color instrument (oci) version 3.1 data products overview. https://pace.oceansciences.org/access_pace_data.htm, 2024. Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) Mission. 2

[5] Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional map of the world. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6172–6180, 2018. 1

[6] Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David Lobell, and Stefano Ermon. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35:197–211, 2022. 1, 2

[7] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 1, 2, 4, 5

[8] Martin Hermann Paul Fuchs and Begüm Demir. Hyspecnet-11k: A large-scale hyperspectral dataset for benchmarking learning-based hyperspectral image compression methods. In *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, pages 1779–1782. IEEE, 2023. 5

[9] Xin Guo, Jiangwei Lao, Bo Dang, Yingying Zhang, Lei Yu, Lixiang Ru, Liheng Zhong, Ziyuan Huang, Kang Wu, Dingxiang Hu, et al. Skysense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27672–27683, 2024. 2

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[11] Yuting He, Fuxiang Huang, Xinrui Jiang, Yuxiang Nie, Minghao Wang, Jiguang Wang, and Hao Chen. Foundation model for advancing healthcare: Challenges, opportunities and future directions. *IEEE Reviews in Biomedical Engineering*, 2024. 1

[12] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019. 1

[13] Oleksii Hrinchuk, Valentin Khrulkov, Leyla Mirvakhabova, Elena Orlova, and Ivan Oseledets. Tensorized embedding layers. In *Findings of the association for computational linguistics: EMNLP 2020*, pages 4847–4860, 2020. 2

[14] He Huang, Quan Wang, Chao Liu, and Chen Zhou. Optimal estimation of cloud properties from thermal infrared observations with a combination of deep learning and radiative transfer simulation. *Atmospheric Measurement Techniques*, 17(24):7129–7141, 2024. 1

[15] Intergovernmental Panel on Climate Change (IPCC). IPCC Official Website, 2024. Accessed: 2024-12-23. 1, 6, 2

[16] Thomas A Jones, David J Stensrud, Patrick Minnis, and Rabindra Palikonda. Evaluation of a forward operator to assimilate cloud water path into wrf-dart. *Monthly weather review*, 141(7):2272–2289, 2013. 6, 2

[17] Hermann Kaufmann, S Förster, Hendrik Wulf, K Segl, Luis Guanter, M Bochow, U Heiden, A Müller, W Heldens, T Schneiderhan, et al. Science plan of the environmental mapping and analysis program (enmap). 2012. 2, 1

[18] Cathy Kessinger, Michael Donovan, Richard Bankert, Earle Williams, Jeffrey Hawkins, Huaqing Cai, Nancy Rehak, Daniel Megenhardt, and Matthias Steiner. Convection diagnosis and nowcasting for oceanic aviation applications. In *Remote Sensing Applications for Aviation Weather Hazard Detection and Decision Support*, pages 77–88. SPIE, 2008. 6, 2

[19] Jingwei Li, Feng Zhang, Wenwen Li, Xuan Tong, BaoXiang Pan, Jun Li, Han Lin, Husi Letu, and Frahan Mustafa. Transfer-learning-based approach to retrieve the cloud properties using diverse remote sensing datasets. *IEEE Transactions on Geoscience and Remote Sensing*, 2023. 1, 2, 6, 8, 4

[20] Jingtao Li, Yingyi Liu, Xinyu Wang, Yunning Peng, Chen Sun, Shaoyu Wang, Zhendong Sun, Tian Ke, Xiao Jiang, Tangwei Lu, et al. Hyperfree: A channel-adaptive and tuning-free foundation model for hyperspectral remote sensing imagery. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 23048–23058, 2025. 2, 8, 3, 4

[21] Sunzhu Li, Peng Zhang, Guobing Gan, Xiuqing Lv, Benyou Wang, Junqiu Wei, and Xin Jiang. Hypoformer: Hybrid decomposition transformer for edge-friendly neural machine translation. In *Proceedings of the 2022 conference on empirical methods in natural language processing*, pages 7056–7068, 2022. 2, 4, 5, 7

[22] Wenwen Li, Feng Zhang, Bin Guo, Haoyang Fu, and Husi Letu. Physics-driven machine learning algorithm facilitates multilayer cloud property retrievals from geostationary passive imager measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–18, 2024. 1

[23] Xuyang Li, Danfeng Hong, and Jocelyn Chanussot. S2mae: A spatial-spectral pretraining foundation model for spectral remote sensing data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24088–24097, 2024. 2

[24] Rebecca Lindsey and David Herring. *MODIS: Moderate Resolution Imaging Spectroradiometer: NASA's Earth Observing System*. Goddard Space Flight Center, 2002. 2

[25] Fan Liu, Delong Chen, Zhangqingyun Guan, Xiaocong Zhou, Jiale Zhu, Qiaolin Ye, Liyong Fu, and Jun Zhou. Remoteclip: A vision language foundation model for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–16, 2024. 1

[26] Peiyu Liu, Ze-Feng Gao, Wayne Xin Zhao, Zhi-Yuan Xie, Zhong-Yi Lu, and Ji-Rong Wen. Enabling lightweight fine-tuning for pre-trained language model compression based on matrix product operators. *arXiv preprint arXiv:2106.02205*, 2021. 2

[27] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 2

[28] Teruyuki Nakajima and Michael D King. Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. part i: Theory. *Journal of Atmospheric Sciences*, 47(15):1878–1893, 1990. 1, 2

[29] NASA Goddard Space Flight Center. PACE Science Data Reprocessing Version 3.x Notes. https://oceancolor.gsfc.nasa.gov/files/data/reprocessing/V3/PACE_Reprocessing_V3.x_notes.pdf, 2025. Accessed: 2025-11-20. 1

[30] Vikas Nataraja, Sebastian Schmidt, Hong Chen, Takanobu Yamaguchi, Jan Kazil, Graham Feingold, Kevin Wolf, and Hironobu Iwabuchi. Segmentation-based multi-pixel cloud optical thickness retrieval using a convolutional neural network. *Atmospheric Measurement Techniques Discussions*, pages 1–34, 2022. 1, 2, 6, 8, 4

[31] Jens Nieke and Michael Rast. Towards the copernicus hyperspectral imaging mission for the environment (chime). In *Igarss 2018-2018 ieee international geoscience and remote sensing symposium*, pages 157–159. IEEE, 2018. 2, 1

[32] Matan Ben Noach and Yoav Goldberg. Compressing pre-trained language models by matrix decomposition. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 884–889, 2020. 2

[33] Mubashir Noman, Muzammal Naseer, Hisham Cholakkal, Rao Muhammad Anwer, Salman Khan, and Fahad Shahbaz Khan. Rethinking transformers pre-training for multi-spectral satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27811–27819, 2024. 1, 2

[34] Rintaro Okamura, Hironobu Iwabuchi, and K Sebastian Schmidt. Feasibility study of multi-pixel retrieval of optical thickness and droplet effective radius of inhomogeneous clouds using deep learning. *Atmospheric Measurement Techniques*, 10(12):4747–4759, 2017. 1, 2, 6

[35] Stefano Pignatti, Angelo Palombo, Simone Pascucci, Filomena Romano, Federico Santini, Tiziana Simoniello, Amato Umberto, Cuomo Vincenzo, Nicola Acito, Marco Diani, et al. The prisma hyperspectral mission: Science activities and opportunities for agriculture and land monitoring. In *2013 IEEE international geoscience and remote sensing symposium-IGARSS*, pages 4558–4561. IEEE, 2013. 2, 1

[36] S Platnick, S Ackerman, M King, K Meyer, WP Menzel, RE Holz, BA Baum, and P Yang. Modis atmosphere l2 cloud product (06_l2), nasa modis adaptive processing system, goddard space flight center. *URL http://dx. doi. org/10.5067/MODIS/MOD06 L*, 2, 2015. 3, 1

[37] S Platnick, KG Meyer, P Hubanks, R Holz, SA Ackerman, and AK Heidinger. Viirs atmosphere l3 cloud properties product. *Version-1.1, NASA Level-1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC), Goddard Space Flight Center*, 2019. 3, 1

[38] CA Poulsen, R Siddans, GE Thomas, AM Sayer, RG Grainger, E Campmany, SM Dean, C Arnold, and PD Watts. Cloud retrievals from satellite data using optimal estimation: evaluation and application to atsr. *Atmospheric Measurement Techniques*, 5(8):1889–1910, 2012. 2

[39] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021. 1

[40] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever.

Zero-shot text-to-image generation. In *International conference on machine learning*, pages 8821–8831. Pmlr, 2021. 1

[41] Colorado J Reed, Ritwik Gupta, Shufan Li, Sarah Brockman, Christopher Funk, Brian Clipp, Kurt Keutzer, Salvatore Candido, Matt Uyttendaele, and Trevor Darrell. Scale-mae: A scale-aware masked autoencoder for multiscale geospatial representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4088–4099, 2023. 2, 5

[42] Linus Scheibenreif, Michael Mommert, and Damian Borth. Masked vision transformers for hyperspectral image classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2166–2176, 2023. 5

[43] Vladan Stojnic and Vladimir Risojevic. Self-supervised learning of remote sensing scene representations using contrastive multiview coding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1182–1191, 2021. 2

[44] Gencer Sumbul, Marcela Charfuelan, Begüm Demir, and Volker Markl. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In *IGARSS 2019-2019 IEEE international geoscience and remote sensing symposium*, pages 5901–5904. IEEE, 2019. 1

[45] Urmish Thakker, Jesse Beu, Dibakar Gope, Ganesh Dasika, and Matthew Mattina. Rank and run-time aware compression of nlp applications. *arXiv preprint arXiv:2010.03193*, 2020. 2

[46] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxvit: Multi-axis vision transformer. In *European conference on computer vision*, pages 459–479. Springer, 2022. 4

[47] Zahid Hassan Tushar, Adeleke Ademakinwa, Jianwu Wang, Zhibo Zhang, and Sanjay Purushotham. Cloudunet: Adapting unet for retrieving cloud properties. In *IGARSS 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 7163–7167. IEEE, 2024. 1, 2, 6, 8, 4

[48] Zahid Hassan Tushar, Adeleke Ademakinwa, Jianwu Wang, Zhibo Zhang, and Sanjay Purushotham. Joint retrieval of cloud properties using attention-based deep learning models. In *IGARSS 2025-2025 IEEE International Geoscience and Remote Sensing Symposium*, pages 4616–4621. IEEE, 2025. 1, 2, 6, 7, 8, 4

[49] Di Wang, Meiqi Hu, Yao Jin, Yuchun Miao, Jiaqi Yang, Yichu Xu, Xiaolei Qin, Jiaqi Ma, Lingyu Sun, Chenxing Li, et al. Hypersigma: Hyperspectral intelligence comprehension foundation model. *PAMI*, 2025. 1, 2, 5, 6, 8, 3, 4

[50] Quan Wang, Chen Zhou, Xiaoyong Zhuge, Chao Liu, Fuzhong Weng, and Minghuai Wang. Retrieval of cloud properties from thermal infrared radiometry using convolutional neural network. *Remote Sensing of Environment*, 278: 113079, 2022. 6

[51] Yue Wang, Ming Wen, Hailiang Zhang, Jinyu Sun, Qiong Yang, Zhimin Zhang, and Hongmei Lu. Hsimae: A unified masked autoencoder with large-scale pre-training for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024. 5

[52] David M Winker, Jacques R Pelon, and M Patrick McCormick. Calipso mission: spaceborne lidar for observation of aerosols and clouds. In *Lidar remote sensing for industry and environment monitoring III*, pages 1–11. SPIE, 2003. 2

[53] Aoran Xiao, Weihao Xuan, Junjue Wang, Jiaxing Huang, Dacheng Tao, Shijian Lu, and Naoto Yokoya. Foundation models for remote sensing and earth observation: A survey. *IEEE Geoscience and Remote Sensing Magazine*, 2025. 2

[54] Shu-wen Yang, Heng-Jui Chang, Zili Huang, Andy T Liu, Cheng-I Lai, Haibin Wu, Jiatong Shi, Xuankai Chang, Hsiang-Sheng Tsai, Wen-Chin Huang, et al. A large-scale evaluation of speech foundation models. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32: 2884–2899, 2024. 1

[55] Maxime Zanella and Ismail Ben Ayed. Low-rank few-shot adaptation of vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1593–1603, 2024. 2

[56] Juanping Zhao, Zenghui Zhang, Wei Yao, Mihai Datcu, Huilin Xiong, and Wenxian Yu. Opensarurban: A sentinel-1 sar image dataset for urban interpretation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:187–203, 2020. 1

[57] Xin Zhou, Yangang Liu, Yunpeng Shan, Satoshi Endo, Yu Xie, and Manajit Sengupta. Influences of cloud microphysics on the components of solar irradiance in the wrf-solar model. *Atmosphere*, 15(1):39, 2023. 6, 2

[58] Yanqi Zhou, Tao Lei, Hanxiao Liu, Nan Du, Yanping Huang, Vincent Zhao, Andrew M Dai, Quoc V Le, James Laudon, et al. Mixture-of-experts with expert choice routing. *Advances in Neural Information Processing Systems*, 35:7103–7114, 2022. 4

[59] Weiming Zhuang, Chen Chen, Zhizhong Li, Sina Sajadmanesh, Jingtao Li, Jiabo Huang, Vikash Sehwag, Vivek Sharma, Hirotaka Shinozaki, Felan Carlo Garcia, et al. Argus: A compact and versatile foundation model for vision. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 4418–4429, 2025. 2