ONBOARD INTELLIGENT SPECTRAL VESSEL DETECTION FOR TIMELY MARITIME SURVEILLANCE

Arthur Vandenhoeke, Olli Eloranta, Susanna Kukkavuori, Michal Shimoni

Kuva Space Oy, Vaisalantie 2-8, 02130 Espoo, Finland

ABSTRACT

Maritime surveillance demands near-real-time intelligence to counter stealth vessels, dark fleets, and AIS spoofing. While hyperspectral imaging provides rich material-specific signatures that complement traditional modalities, the high data volume severely constrains onboard processing and downlink latency. This paper introduces UNetLite3D, a lightweight TensorRT-native 3D U-Net optimized for deployment on NVIDIA Jetson AGX Orin and AGX Thor under strict 10–70 W power budgets. Using FP16 quantization, layer fusion, and Tensor Core acceleration, the model processes full VIS-NIR cubes at high throughput. A key contribution is the domain-specific CopyPasteObjects in-painting augmentation, which samples from a bank of 1 000 pure vessel crops extracted from 600 manually annotated hyperspectral images. On the held-out test set the model achieves IoU = 0.70 and mAP@50 = 0.31. End-to-end benchmarks show that complete 2048×2048 scenes can be processed onboard and packaged into compact GeoJSON reports for satellite-to-satellite downlink within a 15-minute latency target.

***Index Terms*—** Onboard hyperspectral processing, real-time vessel detection, 3D U-Net, TensorRT optimization, maritime domain awareness.

1. INTRODUCTION

The 21st century maritime domain faces escalating threats from stealth vessels, dark-fleet operations, and deliberate Automatic Identification System (AIS) deactivation or spoofing. These tactics facilitate illicit activities such as illegal fishing, sanctions evasion, smuggling, and gray-zone naval maneuvers, significantly degrading traditional Maritime Domain Awareness (MDA) and resulting in response latencies that frequently exceed several hours [1].

Hyperspectral imaging in the Visible-to-Short-Wave Infrared (VIS-SWIR) range offers decisive advantages by capturing rich, material-specific spectral signatures. This enables a complementary vessel detection modality, camouflage discrimination, and classification even under adverse atmospheric conditions such as mist, haze, and sea-state conditions, outperforming conventional multispectral or optical visible sensors. However, the high-dimensional nature of VIS-SWIR data cubes imposes severe challenges on downlink bandwidth, onboard storage, and ground-based processing, making real-time operational use impractical under conventional Earth Observation (EO) architectures [2].

Advances in space-qualified GPU platforms, including NVIDIA® Jetson® AGX Orin™ and AGX Thor™ modules, now enable ultra-low-latency inference directly in orbit. This paradigm shift from raw-data downlink to in-situ intelligence generation transforms EO from a passive observation capability into an active, rapid-response system. The societal impact is substantial: response times are reduced from hours to seconds, directly supporting time-critical applications in maritime security, disaster response, wildfire monitoring, and environmental protection.

Kuva Space's Hyperfield constellation exemplifies this new generation of EO systems. The second-generation microsatellites integrate full VIS-SWIR hyperspectral imaging with high-resolution RGB cameras, AIS receivers, and satellite-to-satellite crosslinks, providing a rich multi-modal data stream for real-time maritime domain awareness.

This work addresses the specific engineering challenge of performing real-time vessel detection from full VIS-SWIR hyperspectral imagery under the stringent power, thermal, and computational constraints inherent to space-borne platforms. The core contributions are threefold: (1) a TensorRT™-optimized UNet architecture designed for efficient onboard inference, (2) a custom in-painting-based augmentation pipeline that significantly improves robustness to spectral variability and the limited labeled data typical of hyperspectral maritime scenes, and (3) comprehensive performance benchmarks including latency, throughput, power consumption, and thermal behavior on flight-representative NVIDIA® Jetson® AGX Orin™ and AGX Thor™ hardware, providing concrete evidence of practical in-situ deployment feasibility for AI-native hyperspectral analytics.

2. RELATED WORK

Hyperspectral imaging has emerged as a powerful solution for detecting and identifying marine vessels, offering strong complementarity to SAR and optical IR/VIS systems. Early approaches relied on classical spectral unmixing, anomaly detection, or edge-based methods. [3] demonstrated aerial

hyperspectral remote sensing for small floating maritime objects using spectral matching and edge detection, yet required extensive offline post-processing. [4] extended similar techniques to UAV hyperspectral imagery for coastline target detection with material-specific spectral libraries. More recent deep-learning works, such as [2], applied CNN classifiers to hyperspectral maritime scenes and reported superior material discrimination over multispectral baselines; however, these approaches remain limited to ground-station processing. Complementary IGARSS 2025 contributions further advance the field: [5] introduced a Constant False Alarm Rate (CFAR) strategy tailored to PRISMA hyperspectral images for robust vessel detection without deep learning, while [6] proposed Vessel-DINO, a multimodal foundation model that fuses optical and additional sensor modalities for improved maritime vessel detection.

UNet-based architectures and their variants have become the de facto standard for semantic segmentation in remote sensing, particularly when dealing with limited labeled data and high-dimensional hyperspectral inputs. Recent hyperspectral-specific advancements include evolutionary hyperparameter optimization of U-Net for enhanced semantic segmentation [7], FFCA-UNet with feature fusion and cross-attention mechanisms [8], DCA-UNet incorporating dual-channel attention to better capture spatial-spectral dependencies [9], and U-Net variants with ASPP and skip connections explicitly designed for hyperspectral images with extreme class imbalance [10]. While these models deliver excellent pixel-wise performance on high-end GPUs, they are almost universally evaluated in full-precision (FP32) without quantization, layer fusion, or hardware-specific optimizations leaving a critical gap for deployment under the severe SWaP (size, weight, and power) constraints of satellite platforms.

Edge-AI optimization techniques on NVIDIA® Jetson® platforms have advanced considerably with TensorRT™ 10.x (current release 10.16 as of April 2026). Layer fusion, FP16/INT8 post-training quantization, dynamic shape support, and Tensor Core acceleration routinely deliver 3–8× reductions in latency and energy consumption compared with native PyTorch® FP32 or ONNX™ Runtime baselines on both AGX Orin™ and AGX Thor™ modules. Independent benchmarks consistently show that non-TensorRT™ implementations suffer from significantly higher power draw and thermal throttling, rendering them unsuitable for sustained operation in spaceborne environments where power budgets are typically limited to tens of watts.

Prior onboard deep-learning demonstrations in spaceborne Earth observation have rapidly evolved from FPGA-centric prototypes to GPU-accelerated commercial-off-the-shelf (COTS) solutions. [11] applied neural architecture search to optimize deep-learning models specifically for on-orbit deployment, achieving sub-megabyte model footprints and real-time inference on Jetson-class hardware while preserving accuracy. In the hyperspectral domain, the CORSA project [12] successfully deployed a unified AI-driven compression and foundation model for hyperspectral data directly on NVIDIA® Jetson Orin NX™ using TensorRT™, demonstrating near-lossless compression ratios up to 300× on EnMAP-like data cubes and enabling downstream analytics in orbit. [13] provide a comprehensive review of deep-learning advancements for onboard hyperspectral processing, highlighting lightweight CNNs and 1D-CNNs as particularly promising for spaceborne constraints. Most recently, Planet Labs [14] executed production-grade AI object detection onboard the Pelican-4 satellite using an embedded NVIDIA® Jetson Orin module, reducing end-to-end insight latency to seconds.

Collectively, these works validate the technical feasibility of in-situ AI for Earth observation. However, none has yet demonstrated a production-grade TensorRT™-optimized UNet for full VIS-SWIR hyperspectral vessel detection under representative maritime conditions, full spectral dimensionality, and flight-representative power/thermal constraints. The proposed work bridges these research strands by delivering exactly such a system—one that outperforms non-optimized and non-hyperspectral baselines in both segmentation/detection accuracy and onboard metrics.

Table 1. UNetLite3D Layer Configuration ($\texttt{base_c} = 24$)

Stage	Channels	Operation
Enc	$1 \to 24$	DC3d
Enc	$24 \to 48$	MP3d + DC3d
Enc	$48 \to 96$	MP3d + DC3d
Bot	$96 \to 192$	DC3d
Dec	$192 \to 96$	CT3d + DC3d
Dec	$96 \to 24$	CT3d + DC3d
Coll	$24 \to 24$	Conv3d ($N_{\text{bands}} \times 1 \times 1$)
Head	$24 \to n_{\text{cls}}$	Conv2d 1×1

DC3d = DoubleConv3d, MP3d = MaxPool3d, CT3d = ConvTranspose3d, Coll = Spectral collapse, n_{cls} = number of classes.

3. METHODOLOGY

To support near-real-time maritime intelligence operations, a satellite mission must process hyperspectral data on board and generate mission-critical insights such as vessel bounding boxes and positions, delivered via traditional S/X-band links or inter-satellite communications, in order to achieve an end-to-end latency of 15 minutes from sensing. Meeting these requirements requires the spaceborne mission to operate under stringent SWaP constraints. For example, the second-generation Hyperfield microsatellites, planned for launch from 2026 onward, allocate only 10–70 W sustained power to the AI compute subsystem (NVIDIA® Jetson® AGX Orin or AGX Thor) while leaving margin for imaging, attitude control, and thermal management. Any viable model must therefore deliver low-latency inference, minimal energy con-

sumption, and full compatibility with TensorRT quantization and layer fusion, all while maintaining high segmentation accuracy on full VIS-SWIR spectra.

Model Architecture. To meet these requirements, we propose **UNetLite3D**—a lightweight, TensorRT-native 3D U-Net specifically engineered for edge deployment on NVIDIA® Jetson® platforms (see Table 1 for the layer configuration with `base_c` $= 24$). The model employs a compact 3D encoder-decoder with `DoubleConv3d` blocks (two 3×3 convolutions + BatchNorm + ReLU, optional residuals) and skip connections to capture spatial-spectral features. An explicit spectral-collapse `Conv3d` layer (kernel size $(N_{\text{bands}}, 1, 1)$) efficiently reduces the spectral dimension in a fully differentiable, TensorRT-compatible manner. With `base_c` $= 24$ the model contains ∼2.8 M parameters; the optional `UNetLite3DMini` variant (`base_c` $= 16$, no residuals) reduces the footprint to ∼1.2 M parameters. Band selection and augmentations are detailed in section 4.

Training and inference pipeline. The model is trained with the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.99$), a cosine learning-rate scheduler with linear warmup (starting at 5×10^{-4}, ramping to 1×10^{-3}, decaying to 1×10^{-6}), and focal cross-entropy loss. Raw output logits are converted to instance masks using `scipy.ndimage.find_objects`. Bounding boxes together with the vessel spectrum are exported into GeoJSON features, packaged onboard into (1) a vessel spectral library and (2) a dark-vessel report (correlated with onboard AIS data). The compact reports are downlinked via satellite-to-satellite crosslinks, achieving the 15-minute end-to-end latency target.

4. DATASET AND AUGMENTATION PIPELINE

The effectiveness of the proposed UNetLite3D architecture is underpinned by a high-quality, domain-specific training dataset and a custom augmentation strategy explicitly designed to address the spectral variability and limited labeled data typical of hyperspectral maritime scenes. We curated a proprietary dataset of 600 manually annotated VIS-SWIR hyperspectral images containing approximately 10 000 vessel instances. The images were acquired by the Hyperfield-1 constellation under diverse atmospheric, sea-state, and illumination conditions. A fixed subset of 11 spectral bands between 490nm and 870nm was selected to balance discriminative power and computational efficiency; these bands are used consistently for both training and inference.

To improve generalization and robustness to rare vessel configurations, we developed a specialized in-painting augmentation based on the `CopyPasteObjects` transform. This mechanism samples from a dedicated bank of 1 000 pure vessel crops (stored as 11-band hyperspectral patches with corresponding binary masks). Figure 1 shows representative RGB previews of the hyperspectral vessel crops stored in the object bank. During training, up to `max_paste` vessels are randomly selected, scaled, rotated, and pasted into valid water regions of the input tile (determined by an NDWI-based thresholded placement mask). Intensity adaptation matches the mean and standard deviation of the pasted object to the local background using buffer-zone statistics, while a cosine-feather edge (3 px) ensures seamless blending. The same geometric and placement operations are deterministically replayed on the corresponding label mask, preserving annotation integrity. Figure 2 illustrates the in-painting process: the top row displays RGB previews, the middle row the updated label mask, and the bottom row the NDWI-derived placement mask guiding valid pasting locations. Additional lightweight transforms (horizontal/vertical flips, 90° rotations, small affine perturbations, spectral jitter, Gaussian noise, and coarse dropout) complete the pipeline, which is applied with probability 0.5.

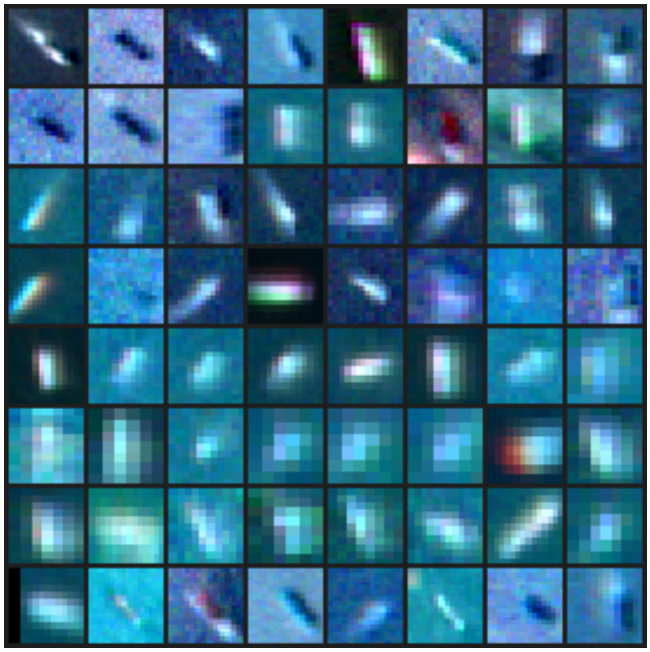


Fig. 1. RGB previews of hyperspectral vessel crops used during in-painting.

To quantify the impact of the in-painting strategy, we evaluate the segmentation performance (IoU and mAP@50) under varying `max_paste` for two model sizes (`base_c`=24 and 32). Performance is measured on the test set that follows a 70/15/15 split (training/validation/test) with scene-level separation to avoid spatial correlation. Results are summarized in table 2. Without in-painting (`max_paste`=0), performance collapses to an IoU of only ∼0.25 and mAP@50 of ∼0.01, revealing severe overfitting and poor generalization to the spectral variability of real maritime scenes. Introducing the CopyPasteObjects augmentation yields dramatic gains. For `base_c`=24, `max_paste`=6 achieves the highest IoU (0.701) with strong mAP@50 (0.284). For

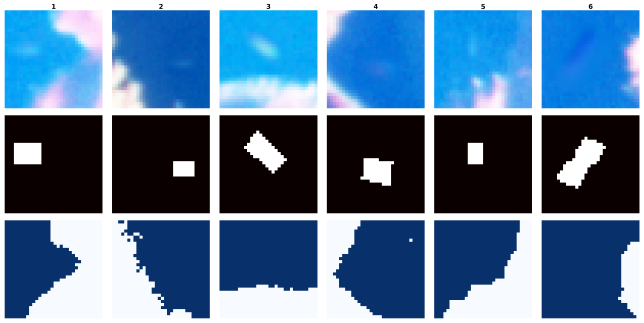


Fig. 2. Examples of in-painting a vessel from the ship bank into a hyperspectral tile. Top row: RGB previews. Middle row: associated label mask. Bottom row: approximate NDWI mask where vessels are allowed to be in-painted.

Table 2. Impact of in-painting augmentation (`max_paste`) and model size (`base_c`) on segmentation performance.

base_c	max_paste	IoU	mAP@50
24	0	0.248	0.014
24	3	0.655	0.284
24	6	**0.701**	0.284
24	9	0.667	0.270
32	0	0.247	0.009
32	3	0.676	0.266
32	6	0.679	**0.310**

`base_c`=32, `max_paste`=6 provides the best overall balance (IoU = 0.679, mAP@50 = 0.310). Higher paste counts (`max_paste`=9) slightly degrade mAP@50, indicating that excessive augmentation can introduce unrealistic crowding. These results demonstrate that the proposed domain-specific in-painting augmentation is by far the dominant contributor to model performance and is essential for robust hyperspectral vessel segmentation under real-world maritime conditions. Nevertheless, the technique also has limitations. Because vessel signatures are blended with local background statistics, the generated samples can contain non-physical spectral composites. Consequently, there is a risk that the model may learn artifacts that do not exist in real maritime scenes. While the complementary geometric, radiometric, and illumination transforms help reduce overfitting to the specific vessel crops, they do not completely eliminate this issue.

5. ONBOARD MODEL BENCHMARK

To validate the feasibility of real-time onboard hyperspectral scene processing, we evaluate the TensorRT-optimized inference pipeline on two flight-representative NVIDIA® Jetson® platforms: the AGX Orin™ and the AGX Thor™. The goal is to prove that the optimized pipeline can process a complete 2048×2048 pixel maritime scene for vessel detection under the 10–70 W power envelope allocated to the AI compute system on the Hyperfield microsatellites. All benchmarks employ the UNetLite-3D `base_c` = 24 model variant with 11 input bands, compiled as optimized TensorRT™ model engines.

Table 3. The Jetson device benchmark results on the UNetLite-3D (base_c $= 24$) model.

Parameters				TensorRT (tile-level)			E2E
Mode (W)	Prec.	Tile (px)	Best BS	Lat. (ms)	Energy (J/img)	Eff. (GFLOP/J)	Prod. /hr
AGX Orin							
35	FP16	64	4	10,260	239	74.1	221
35	FP16	128	2	10,549	258	68.6	228
35	FP16	256	1	10,471	262	67.5	223
35	FP32	64	1	72,550	941	18.8	45
60	FP16	64	4	4,946	237	75.0	329
60	FP16	128	2	5,215	261	68.0	287
60	FP16	256	1	5,312	274	68.6	267
60	FP32	64	1	28,569	578	30.6	103
AGX Thor							
60	FP16	64	4	4,311	198	89.7	325
60	FP16	128	1	4,319	200	88.6	403
60	FP16	256	1	4,315	203	87.3	402
60	FP32	64	4	15,647	603	29.4	168
90	FP16	64	4	3,409	205	86.1	454
90	FP16	128	2	3,395	213	83.1	522
90	FP16	256	1	3,328	217	81.8	520
90	FP32	64	4	12,441	630	28.1	277
120	FP16	64	4	2,290	**185**	**98.3**	568
120	FP16	128	4	2,291	189	96.1	666
120	FP16	256	2	**2,280**	193	94.2	**679**
120	FP32	64	4	8,345	547	32.4	303

Both platforms expose configurable power modes governing maximum CPU/GPU clock frequencies and active core count. We benchmark the AGX Orin at 35 W and 60 W profiles. For the AGX Thor, a matching custom 60 W mode enables direct cross-platform comparison, while additional 90 W and 120 W modes explore the throughput gains available beyond the nominal satellite power envelope for missions that afford higher power consumption.

The benchmark is structured in two stages. At the **tile level**, we profile individual inference calls across combinations of power mode, model precision (FP32, FP16), and square window sizes. The results are calculated over a benchmarking period of 30 s. The reported tile-level metrics are normalized to full 2048×2048 scene equivalents.

The **end-to-end** pipeline additionally includes data loading, tiling, batched inference, and reassembly, replicating the complete inference service as deployed onboard the satellite. The resulting metric, *products per hour*, quantifies how many complete 2048×2048 scenes can be processed per hour and serves as the primary indicator of operational throughput. This is calculated as an average over 5 batches of images. For both benchmarks, the optimal batch size is chosen, and for the end-to-end pipeline the optimal number of data-loading

workers is jointly optimized to keep the GPU fully occupied while hiding I/O latency.

The benchmark results are presented in table 3. Both the AGX Orin and AGX Thor achieve over 300 products per hour at the 35–60 W power modes that fall within the Hyperfield satellite's operational envelope, confirming that real-time onboard vessel detection is feasible under representative SWaP constraints. We confirm the importance of FP16 quantization, as it performs at a 3–4 times higher efficiency than FP32, depending on available power. At the common power mode, the AGX Thor delivers an average 26% improvement in power efficiency (GFLOP/J) and 30% in processed products per hour relative to the AGX Orin.The Thor performs more efficiently on larger tile sizes, which is notable as larger tiles provide more spatial context and therefore improve model performance. When the higher power modes are engaged, the AGX Thor extends its advantage to up to 106% higher throughput and 31% higher power efficiency than the best Orin configuration, demonstrating that the additional power headroom translates into increasingly efficient throughput gains.

6. CONCLUSION

This work presented UNetLite3D, a lightweight TensorRT-native 3D U-Net for real-time onboard hyperspectral vessel detection. Combining a compact 3D encoder-decoder with an explicit spectral-collapse layer and a domain-specific Copy-PasteObjects in-painting augmentation, the model achieves IoU = 0.70 and mAP@50 = 0.31 on a test set of 600 annotated VIS-SWIR scenes containing approximately 10 000 vessels. Benchmarks on NVIDIA® Jetson® AGX Orin™ and AGX Thor™ confirm that the FP16-optimized engine meets the 10–70 W power envelope required for operational use, with capability to process over 300 hyperspectral products per hour.

These results increase confidence that, through Hyperfield-2's in-orbit integration of full VIS-SWIR hyperspectral imaging, VHR optical cameras, AIS receivers, and inter-satellite crosslinks, the system will be capable of end-to-end onboard intelligence generation and delivery of dark-vessel alerts within the targeted 15-minute latency. Future work will explore multi-modal fusion and bitemporal change detection.

7. REFERENCES

[1] M. Kastler, "Move beyond ais for maritime domain awareness," 2025, Proceedings of the U.S. Naval Institute.

[2] D. Seo, D. Lee, S. Park, and S. Oh, "Hyperspectral image-based identification of maritime objects using convolutional neural networks and classifier models," *Journal of Marine Science and Engineering*, vol. 13, no. 1, pp. 6, 2025.

[3] J. J. Park and K. A. Park, "Aerial hyperspectral remote sensing detection for maritime search and surveillance of floating small objects," *Advances in Space Research*, 2023.

[4] S. Zhao, Y. Lv, X. Zhao, J. Wang, and W. Li, "Coastline target detection based on uav hyperspectral remote sensing images," *Frontiers in Marine Science*, vol. 11, 2024.

[5] S. Maresca and S. Matteoli, "Cfar vessel detection in prisma hyperspectral images," in *2025 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, August 2025.

[6] C. Li, D. Hong, B. Zhang, and J. Chanussot, "Vesseldino: Multimodal foundation models for maritime vessel detection," in *2025 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2025.

[7] Laritza Pérez-Enríquez, Saúl Zapotecas-Martínez, Leopoldo Altamirano-Robles, Raquel Díaz-Hernández, and José de Jesús Velázquez Arreola, "Evolutionary optimization of u-net hyperparameters for enhanced semantic segmentation in remote sensing imagery," *Earth*, vol. 7, no. 2, pp. 34, 2026.

[8] Libin Chen, Zihan Li, Xiongwu Xiao, Mohamed M. Elisy, Weiwei Wu, and Deren Li, "Ffca-unet: Feature fusion and cross-attention unet for remote sensing semantic segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 19, pp. 3908–3923, 2026.

[9] Kunbo Han, Mingjin Chen, Chongzhi Gao, and Chunmei Qing, "Dca-unet: Enhancing small object segmentation in hyperspectral images with dual channel attention unet," *Journal of the Franklin Institute*, vol. 362, no. 4, 2025.

[10] Nian Zhang and Tewelde Semere, "U-net based semantic segmentation of hyperspectral images with extreme class imbalance," *IEEE International Conference on Image Processing (ICIP)*, pp. 27–32, 2026.

[11] Roberto del Prete, Parampuneet Kaur Thind, Andrea Mazzeo, Matthew Whitley, Lorenzo Papa, Nicolas Longépé, and Gabriele Meoni, "Optimizing deep learning models for on-orbit deployment through neural architecture search," *Scientific Reports*, vol. 15, pp. 37783, 2025.

[12] X. Ivashkovych, N. Witvrouwen, B. Beusen, A. Luytsand, and T. Van Achteren, "Implementation of corsa on nvidia jetson hardware," 2024, VITO Remote Sensing / ESA Φ-Lab.

[13] Nafiseh Ghasemi, Jon Álvarez Justo, Marco Celesti, Laurent Despoisse, and Jens Nieke, "Onboard processing of hyperspectral imagery: Deep learning advancements, methodologies, challenges, and emerging trends," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 4780–4790, 2024.

[14] Planet Labs PBC, "Planet successfully runs ai in space," Business Wire / Satellite Today, April 2026.