

HyperspectralViTs: General Hyperspectral Models for On-Board Remote Sensing

Vít Růžička  and Andrew Markham 

Abstract—On-board processing of hyperspectral data with machine learning models would enable an unprecedented amount of autonomy across a wide range of tasks allowing new capabilities, such as early warning systems and automated scheduling across constellations, of satellites. However, current classical methods suffer from high false positive rates and therefore prevent easy automation while previously published deep learning models exhibit prohibitive computational requirements. We propose fast and accurate machine learning architectures, which support end-to-end processing of data with high spectral dimension without relying on handcrafted products or spectral band compression techniques. We create three new large datasets of hyperspectral data containing all relevant spectral bands from the near global sensor EMIT. We evaluate our models on two tasks related to hyperspectral data processing—methane detection and mineral identification. Our models reach a new state-of-the-art performance on the task of methane detection, where we improve the F1 score of previous deep learning models by 27% on a newly created synthetic dataset and by 13% on the previously released large benchmark dataset. Our models generalize from synthetic datasets to data with real methane leak events and boost performance by 6.9% in F1 score in contrast with training models from scratch on the real data. Finally, with our newly proposed architectures, one capture from the EMIT sensor can be processed within 30 s on a realistic proxy of the ION-SCV 004 satellite and in less than 0.64 s on a GPU powered Jetson AGX Xavier board.

Index Terms—Efficient machine learning, hyperspectral machine learning, machine learning for imaging spectroscopy, methane detection, mineral identification, on-board deployment.

I. INTRODUCTION

IN this article, we introduce HyperspectralViTs, new adaptations of Transformer-based machine learning models for semantic segmentation of hyperspectral data (also referred to as imaging spectroscopy data) for low compute environments. We demonstrate these adaptations on two recent architectures, SegFormer [1] and EfficientViT [2] and for two relevant tasks, methane leak detection and mineral identification. We describe the existing landscape of approaches used to detect methane leaks and minerals in remote sensing data and outline several limitations to these methods. We then use these as motivation to propose changes to the Transformer-based architectures. We

Received 24 October 2024; revised 9 January 2025; accepted 29 March 2025. Date of publication 3 April 2025; date of current version 21 April 2025. (Corresponding author: Vít Růžička.)

The authors are with the University of Oxford, OX13QD Oxford, U.K. (e-mail: vit.ruzicka@cs.ox.ac.uk; andrew.markham@cs.ox.ac.uk).

Code and data are available online at <https://github.com/previtus/HyperspectralViTs>.

Digital Object Identifier 10.1109/JSTARS.2025.3557527

improve the accuracy and inference speed of the existing methods and benchmark the model in limited compute environment, which serves as a proxy to the hardware available on the satellite.

Methane detection is an important task, which could lead to the reduction of anthropogenic methane in atmosphere—this has been identified as one of the fastest pathways to reduce global warming in United Nations Global Methane Assessment report [3]. In practice, the detection of methane leaks remains to be manual, as the outputs of existing methods can be quite noisy. For large scale automation of this process, reliable and accurate models are needed to sift through the increasing deluge of data from newly deployed hyperspectral sensors.

Mineral identification is a task related to scene decomposition into individual endmembers from a spectral library. On its own it could serve as a strong general task highly relevant for future Foundation models working with hyperspectral data. Hyperspectral Foundation models have been explored in the very recent works of [4] and [5]. We note that one of the example downstream tasks of interest could also be methane detection, as it has been shown that having information about scene background is beneficial for separation of methane leak events from confounders [6].

Autonomy on-board of satellites is needed to enable rapid response, alerting and automatic scheduling within a constellation of satellites [7], [8]. Recent works have shown that detection of events of interest (such as floods) can be done directly in space with on-board machine learning models, reducing the quantity of data for downlinking [9].

We note that the proposed architecture is aimed at the general problem of processing hyperspectral data and as such can be used outside of the domain of remote sensing. By framing the problem as an end-to-end semantic segmentation, we disentangle the proposed architecture from the task of methane leak detection and retain generality.

Developing machine learning architecture for hyperspectral data are well timed and necessary for many downstream applications. Hyperspectral remote sensing data are used in mineralogy [10] and agriculture [11]. Sensors with high spectral dimension are deployed on devices used for deep space exploration [12], [13], [14] and on hand-held devices [15] including mobile phones [16].

Our main contribution is the proposal of a new family of models called HyperspectralViTs, architectures adapted to handle hyperspectral data without informational losses due to reliance on classically computed intermediate products, or due to informational bottlenecks in the original architectures as illustrated

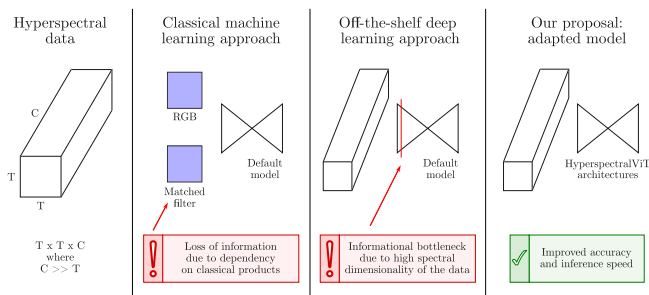


Fig. 1. Illustration of the limitations of previous approaches. First, dependence on classically used products computed from hyperspectral data (such as matched filters in [17]) may lead to loss of accuracy due to imperfect capture of the events of interest. Second, off-the-shelf versions of machine learning models are not adapted to hyperspectral data and may lose valuable information in the early layers of the models. Finally, our adapted model leverages information from all relevant hyperspectral bands, which leads to improvements in both accuracy and inference speeds of our method.

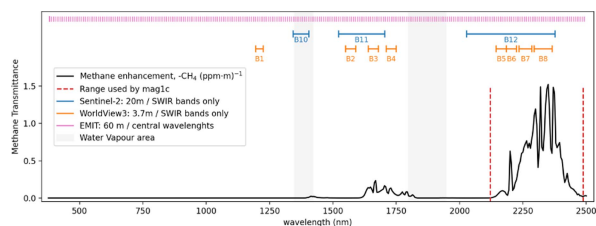


Fig. 2. Methane gas signature (shown through the methane transmittance) in comparison with the band ranges of typically used multispectral and hyperspectral satellites.

on Fig. 1. We analyze the limitations of existing machine learning architectures without adaptations for hyperspectral data and show that we can improve their performance with several simple steps.

Namely, we propose the HyperSegFormer and HyperEfficientViT variants of the default models and evaluate them on a newly created dataset of synthetic methane leak events in the data from the near global hyperspectral sensor EMIT [18]. We then demonstrate that these gains translate over to a newly created dataset of real methane leak events from the same sensor. We show that we are able to improve the accuracy and the inference speed of the existing classical and machine learning approaches by removing their dependency on the classical matched filter product.

Finally, we demonstrate that our models are relevant also for other tasks of interest on a newly created dataset for mineral identification.

A. Literature Review

In this article, we primarily tackle the tasks of methane leak detection, as such most of the relevant literature focuses on this discipline. We however also keep a brief section on methods used for mineral identification.

The task of methane leak detection can be more generally described as semantic segmentation of remote sensing images and has recently seen increased interest from the machine learning community [19].

Fig. 2 shows the wavelength range of typically used multispectral and hyperspectral satellites alongside with the target signature of methane (modelled from the HITRAN 2012 database [20]). The multispectral satellite Sentinel-2 has been previously used to monitor large methane leaks in [21], however multitemporal passes over the same location were needed to detect weaker events. Sánchez-García et al. [22] used data from the WorldView-3 satellite, which has multiple bands in the desired area of interest, but only large methane leak events can be detected and validation requires human expert intervention.

Hyperspectral data provides more information for the task of detecting methane signal, this is due to the improved sampling in spectral ranges where methane is visible (namely, between 1600–1800 nm and 2100–2500 nm). This task has been classically addressed with matched filter approaches [23] which compare the observed data against the target signature of methane. The resulting matched filter product, sometimes called “methane enhancement,” is quite noisy, as other confounder materials (for example rooftops, roads, and solar panels) get detected as well. Iterative matched filter approaches, such as the mag1c filter [24] attempt to improve the clarity of the product. Other works depend on matched filter variants using wider spectral windows [25]. Up until recently, these methane enhancement products were manually inspected, and trained experts were tasked to create final detections.

Mineral detection in remote sensing data has been approached with a wide range of classical methods, which are described in the overview of [26], and only a subset of these uses hyperspectral data. Mineral identification in hyperspectral data can be formalized as spectral unmixing, that is identifying which individual endmembers from a spectral library of chemical and mineral constituents create the observed signature. One classical approach is the Tetracoder program [27] that uses a cascade of manually derived rules to compare observed spectral feature with a reference spectral library of minerals. It isolates discriminative portions of spectra known for each mineral using continuum removal techniques [28] and then computes degree of fit metrics for each item in the library. It has been described as a simulation of the process a trained mineralogist would do when looking at the hyperspectral data. This task can also be formalized as semantic segmentation, using “multihot” labels (multiple minerals can be present in each pixel).

We note that datasets for global mineral identification across the entire globe are missing. Works, such as [29], evaluate their methods on just two airborne hyperspectral captures. Works, such as [27], compare their methods against several well studied scenes such as the Cuprite or Alunite Hill scenes—these have been sparsely in-situ sampled to provide verification data. Note that large scale differences exist when comparing in-situ measurements and space based instruments (where one pixel can cover over e.g., 60×60 m area). Therefore, existing works mostly remain local in their scope, and the used datasets are quite small. Machine learning models adapted to handle hyperspectral data, such as [30], [31], were typically trained only on very small datasets. With our work we aim to train our proposed models on much larger and more diverse datasets, while using multiple tasks.

As has been identified in [32] and several overview papers [33], [34], [35], research in machine learning for processing hyperspectral data has been limited mainly due to the lack of large, publicly available datasets with high quality annotations. Many works still use very small datasets, such as Indian Pines [36], University of Pavia, Houston, or Salinas, all of which contain only a few data captures. As further noted in [33], this data is often divided into almost equal train and test subsets, which leads to model overfitting and results with very high accuracies almost regardless of the used model. In the domain of methane leak detection, only the recent works of [6], [17] release datasets created from a larger archive of full hyperspectral data captures provided by NASA. Růžička et al. [17] furthermore refines the available labels. The authors in [37] and [38] proposed machine learning models for detection and quantification of methane leak events in hyperspectral data, however the used datasets and code/model weights are not published. In contrast, we publish all our used code and trained models and present the created datasets in a machine learning ready format.

Contribution of our work compared to the prior works is the focus on the low compute environment and the corresponding need for highly efficient models. We propose our models with this constraint in mind and we also test them on limited compute devices that serve as proxies to the computational environment on-board of satellites [7], [39]. We note that one could use off-the-shelf efficient machine learning models [2] or apply methods, such as pruning [40], to existing machine learning architectures to improve the inference speed of the system. However, this would not address the problem of informational loss either due to the reliance on classical products, or due to informational bottlenecks encountered in not adapted architectures.

Deployment of machine learning models on-board of satellites with hyperspectral sensors was shown in [41] to filter cloudy data captures. The demonstrator Earth Observer 1 (EO-1) mission by NASA hosted simple thresholding models that processed hyperspectral data on-board to detect clouds and preclassify scenes [42] and later used models for mineral identification in [43]. On-board deployment of unsupervised models to detect anomalous events in multispectral data was proposed in [44] and later tested on-board of a real satellite in [45]. These examples motivate us that methane detection using deep learning models on the current generation of hardware on today’s satellites would be feasible. They also highlight the potential benefits from autonomy on-board in future satellite constellations.

II. DATA

In order to train machine learning models, which would be able to generalize globally around the world, we primarily need global and large datasets with high quality labels. As we already noted, many of the typically used hyperspectral datasets are very small. Recent methane leak datasets from [6], [17] are large enough to train deep learning models, however they both use data from the aerial AVIRIS-NG sensor, which is relatively local in scope (e.g., the Four Corners region of USA). Instead, we use the data from the recent hyperspectral sensor Earth Surface

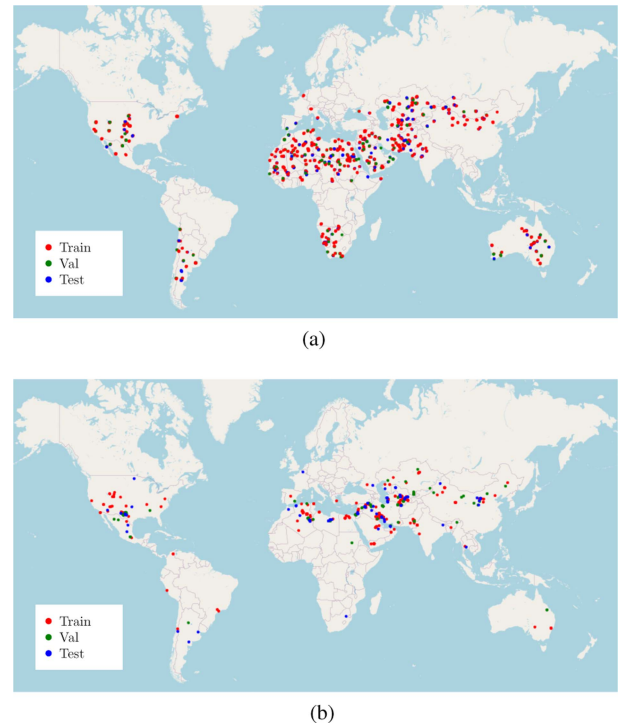


Fig. 3. Locations of tiles used to create the OxHyper datasets of EMIT data. We note that the same source tiles (including the same dataset splits) are used for the minerals (where they contain more spectral bands) and the synthetic methane datasets (where the methane leak events are added into the clean datacubes). (a) OxHyperSyntheticCH4, OxHyperMinerals. (b) OxHyperRealCH4.

TABLE I
OVERVIEW OF DATASETS CREATED IN THIS ARTICLE

OxHyper-	Bands	Spectral range (nm)	Train	Val	Test
SyntheticCH4	86	1573–1699, 2004–2478	796	198	200
RealCH4	86	1573–1699, 2004–2478	279	91	98
Minerals	285	381–2492	796	198	200

STAR COP [17] 125 RGB_c 1573–1699, 2004–2480 3425 N/A 342

We report the number of large tiles of 512×512 Px, note that we usually further tile these into 64×64px tiles with 32px overlap.

Mineral Dust Source Investigation (EMIT), which is deployed on the International Space Station and has near global coverage of the arid regions of the Earth [18]. In total, we create three new datasets from EMIT data, their global coverage is demonstrated in Fig. 3 and they are described in detail in Table I.

A. Methane Leak Event Simulation

In one of our datasets, we simulate synthetic methane leak events in the clean hyperspectral datacubes, given the knowledge of the expected signal of methane and plume concentration data from real events. Unlike the previous study of [46], which used plume shapes simulated using the expansive Large Eddy Simulation [47], we use real methane leak events from the similar hyperspectral sensor AVIRIS-NG, to inform the simulation. Namely, we use the STAR COP dataset [17], given it is improved and manually refined labels and high diversity of plume sizes.

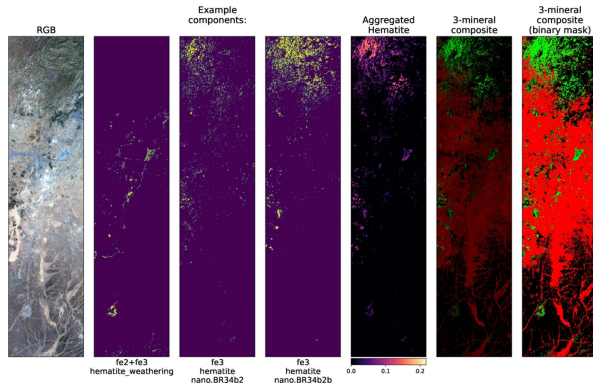


Fig. 4. Aggregation of mineral components into mineral classes. From left to right we show a RGB visualization of the scene, binary mask of three common components for Hematite, the aggregated Hematite product (using in total 14 components) and visualizations of a 3-minerals composites. We show the “Goethite,” “Hematite,” and “Kaolinite,” composite and the binary maps used for training (in the last column).

Mathematically, we can describe the relation between a datacube with and without methane leak event using the Beer-Lambert absorption law [24] as

$$L_{\text{simulated}}(\alpha, s) = L_{\text{clean}} * e^{-\alpha * s} \quad (1)$$

where L refers to at sensor radiances in the hyperspectral datacubes (clean and resulting simulated one), α to the concentration of methane at given location and s the methane absorption at given wavelength. We note that the quantity α is usually the one we are trying to estimate with matched filter approaches [24]. As such, we can use existing matched filter products as sources for this quantity when simulating synthetic events. For s we query the values of the methane signature at the wavelength of each band as plotted on Fig. 2.

We then recompute the matched filter product from the simulated datacubes, this product now contains both the simulated event and the background profile of the scene (including confounder noise).

B. Mineral Aggregation

To create our new mineral dataset, we adapt the steps described in the Theoretical basis document [48] released alongside the “EMITL2BMIN.001” product [49] computed from the EMIT data. This data contains identifications of 381 individual components (separated into two groups). These can be aggregated into larger mineral classes, for this we select the so-called “EMIT10” subset defined in [48]. The aggregation process is shown on Fig. 4. We note that we use only the binary maps of three selected mineral classes out of the total 10. We chose the subset $M = \{\text{Goethite, Hematite, Kaolinite}\}$, as these classes were abundant in the locations sampled in our dataset. We use the released annotations which were computed using the Tetracorder method [27], and as such we cannot really use Tetracorder as a classical baseline (as it would unfairly obtain perfect scores). Therefore, we consider the created dataset as having pseudoground truth labels, and our results only as initial proof-of-principle. We however note, that given any manually validated dataset, our models could be easily retrained with the new data.

C. Final Created Datasets

First, we created the “OxHyperSyntheticCH4” dataset of synthetically simulated methane leak events using the labels of the STARCOP dataset with newly downloaded near-global hyperspectral datacubes from the EMIT sensor (in L1B level of data processing). In total we keep 796 large tiles (of 512×512 px) in the training dataset, 198 in the validation dataset and 200 in the test dataset. Each subset is made from completely nonoverlapping sources of data (clean EMIT datacubes and labels from the STARCOP dataset from AVIRIS-NG data). In each subset we uniformly sample the available plume sizes, so that we have similar distribution of event sizes. Half of these tiles contain simulated events while the other half is kept event free to provide negative samples. We note that these large tiles (of 512×512 px) are further tiled when training the models. All datasets created in this article use tile size of 64 pixels with an overlap of 32 pixels. When evaluating the models (either on the Val or Test subsets), we keep the full resolution of 512×512 px. The synthetic train dataset gets tiled into 174 341 samples. The total size of this data is 228 GB.

Second, we create the “OxHyperRealCH4” dataset of real methane leak events. In total we have 279 large tiles in the training dataset, 91 in the validation dataset and 98 in the test dataset. When further tiled, the training dataset has 54 805 samples. These have reliable ground truth labels, which we calculated from the original labels provided on the EMIT data portal and consequentially manually checked. We note that this dataset is smaller than the other ones and training a machine learning model from scratch could lead to overfitting and poor generalization, however it can also be used for model fine-tuning and to test generalization ability of models trained on other datasets. The total size of this data is 47 GB.

Third, we create “OxHyperMinerals,” a completely new dataset for mineral identification, with labels of 3 selected mineral classes and source annotation for all 381 constituents (separated into two groups). In total we have 796 large tiles in the training dataset, 198 in the validation and finally 200 in the test dataset. When further tiled, the training dataset has 174 341 samples. The large tiles exactly match the tiles used in the synthetic methane events dataset, however they do not contain the simulated events and contain more bands. The mineral dataset contains all 285 spectral bands in between spectral range of 381–2492 nm. The total size of this data is 372 GB.

Finally, we also use a variant of the STARCOP dataset from [17] with all relevant spectral bands extracted from the source data. The STARCOP dataset uses tile size of 128 pixels with an overlap of 64 pixels for its Train subset. The Test subset is kept in its original resolution of 512×512 px. We train our models and compare them on this benchmark dataset with the previously reported results.

III. METHODOLOGY

A. Limitations of Existing Approaches

We are using three different semantic segmentation architectures with the created dataset of hyperspectral datacubes—the commonly used U-Net model [50] and the newer SegFormer [1]

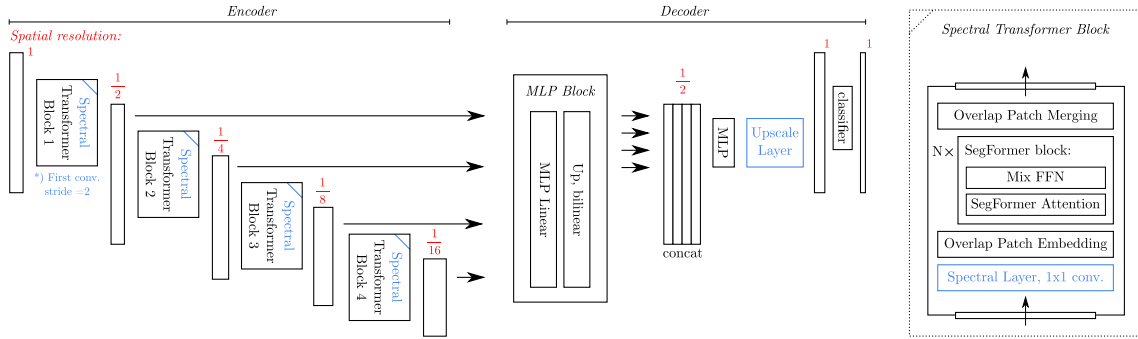


Fig. 5. Illustration of the Hyper SegFormer model for semantic segmentation of methane plumes in Hyperspectral data. We highlight in blue the three proposed modular adjustments. 1) Spectral layers (denoted as “ 1×1 Conv”) in the Transformer blocks, 2) Upscaling layer in the decoder network (denoted as “Upscale layer”), and finally 3) adjusting the stride of the first Transformer block to 2 (denoted as “Stride”). In red, we show an example of the progressive decimation of the resolution throughout the model (using all three adjustments)—we highlight that the typical off-the-shelf variants of the models reduce the output resolution by a factor of 4.

and EfficientViT [2] models. These models can be understood in terms of the encoder–decoder layout. The U-Net model consists of downsampling convolutional layers in the encoder network, and upsampling layers in the decoder, while also using skip-connections to communicate across the same spatial resolution of the data. The SegFormer model uses Transformer blocks in its encoder network, concatenates the features from different resolution levels and then uses a MLP block in the decoder network. The EfficientViT model further proposes changes to make the model run faster on limited hardware. In all models the spatial resolution is initially reduced and then reconstructed in the decoder, while information at different scales is kept with the use of skip-connections, or when concatenating the features together.

We are interested in training these architectures in an end-to-end manner on the full ranges of selected bands of hyperspectral datacubes. In the methane detection task, this allows us to not depend on the classical matched filter product, which can be slow and which contains false detections of confounder materials. Models designed with typical computer vision data in mind (with one to three bands) however suffer from what we call informational bottleneck when they are used with data with large spectral dimensionality. As an example, if we select 86 bands of data which are relevant for our task of methane detection, the first layer in the vanilla versions of the used models reduces the data volume by -95.34% in the case of either the SegFormer, EfficientViT, or U-Net (instead of increasing this volume by $+33\%$ as is common with RGB data). We claim that this informational bottleneck forces the model to get rid of most of the potentially valuable information right at the entry point to the model, which means that the later layers cannot effectively extract high level compressed information, as this extraction has to occur in the first convolutional layer.

Furthermore, we note that the SegFormer and EfficientViT architectures reduce the spatial resolution of the input data by the factor of 4, and use nonparametric bilinear interpolation to scale the predicted output to the original resolution. In contrast, the U-Net architecture reduces the spatial dimension of the data only by the factor of 2, and uses a learned Upscaling layer at the end of the model to reconstruct the original resolution.

B. Proposed Architecture

We propose three modular changes to the selected architectures to allow for better handling of data with high spectral dimension (such as data from hyperspectral satellites) as illustrated on Fig. 5.

First, to aid with the spectral dimensionality of the data, we propose an addition of a special spectral layer into all Transformer blocks throughout the encoder networks. We add a convolutional layer with kernel size of $(1,1)$, stride of 1 and padding of 0, keeping the output of each layer with the same spectral dimension as on its input. This allows for learning operations that occur merely on the spectral data, on its own these allude to pixel-by-pixel operations of classical methods. While in theory the original convolutional layers with larger kernel sizes could learn similar operations, by restricting the operation to just the 1×1 convolutions, we obtain much sharper outputs. Effectively, this also allows the model to weight different input bands with a mechanism similar to attention, thus selecting only the relevant information for the task of interest. While this is a simple addition to the model, we note that it is quite effective, and it does not significantly increase the computational requirements of the model.

Second, we address the reduction of the spatial resolution of the data. We implement a similar Upscaling layer to the one used in the U-Net architecture at the end of the Transformer-based models, just before their classification heads. This block is composed out of two repeated 2-D convolutional layers with kernel size of $(3,3)$, padding and stride of 1, each followed by a ReLU activation layer and Batch normalization layer.

As an additional solution for the recovery of the lost resolution, we also explore the change to the hyperparameters of the original architectures, namely, changing the stride of the first convolutional layer (e.g., in SegFormer from 4 to 2). This change effectively increases the spatial size reduction of the data propagated through the network by the factor of 2. This does not change the number of parameters of the resulting model, but as the effect of this change influences the whole model, the impact on the inference speed is larger than with the Upscaling layer.

These three proposed changes are modular and can be evaluated in isolation or used all together—the resolution of the

output product will either be $\frac{1}{4}$, $\frac{1}{2}$, or 1 times the resolution of the input.

C. Baselines

As a classical baseline approach for methane detection in hyperspectral data, we use the matched filter product (using the iterative variant `mag1c` [24]) with the threshold of 500 ppm \times m. In addition, we apply an opening morphological filter to remove isolated noisy pixels in the product. More specifically, we perform the erosion and dilation operations (using the implementations from the Kornia library). We explore two kernel shapes: “cross” kernel (3×3 pixels matrix with the middle row and column full of ones) and “ones” kernel (3×3 pixels matrix with all values set to ones). We observed small differences when trying these kernels, the “cross” kernel further corresponds to the method used in [17]. We observed that other kernel shapes, larger sizes, or different initial thresholds did not improve the results.

As a classical machine learning baseline, we implement the HyperSTARCOP model based on the U-Net architecture proposed in [17] and train it on our new methane datasets. This model uses the precomputed matched filter product alongside the RGB bands of the input data and effectively learns to clean the classical methane enhancement product. We note, that if the matched filter product misses a methane leak event, this model has no way of accessing the original information in the hyperspectral datacube to recover the detection.

For the task of mineral identification, we cannot use the classical method of Tetracorder as a baseline, as we used it to generate our pseudoground truth labels. When inspecting other baseline approaches such as for example the spectral angle mapper [51], we noted significant disagreement between the two methods. As such, for the task of mineral identification, we chose to only compare different versions of our adapted models against their default out-of-the-shelf version which we treat as a deep learning baseline.

IV. EXPERIMENTAL SETUP

When training machine learning models on our newly created datasets, we use the following settings and hyperparameters. We train our models on the task of semantic segmentation, in methane detection with two exclusive classes (“plume” and “no plume”) and in mineral detection with a multihot label (which allows no detection or potentially multiple active classes of “Goethite,” “Hematite,” and “Kaolinite”).

During training on the methane datasets, to balance the data distribution, we oversample the instances with plume pixels by using the `WeightedRandomSampler` provided by the PyTorch library. We use the Adam optimizer with learning rate of 0.001 for the U-Net architectures and 0.00006 for the SegFormer and EfficientViT architectures. U-Net models use the MobileNet-v2 backbone [52] in the encoder, SegFormer uses the B0 backbone architecture and EfficientViT the B1 backbone architecture. We use the batch size of 32 with U-Net models and 16 with SegFormer models, due to memory limitations while training.

On the “OxHyperSyntheticCH4” dataset we train all models for 50 epochs. When training we use the loss on the set aside validation subset to select the best performing models. When fine-tuning the models on the “OxHyperRealCH4” dataset, we train the models for further 10 epochs from the best reached checkpoint. When training models from scratch on this smaller dataset, we keep the number of epochs to 50. End-to-end models use 86 bands of the EMIT data (RGB and bands between 1573–1699 nm and 2004–2478 nm).

To compare the newly proposed models with the HyperSTARCOP “RGB+MF” model from [17], we also train them on the AVIRIS-NG STARCOP dataset. Here, we train the models for 15 epochs using a batch size of 16 and tile size of 128 px with an overlap of 64 px. Except for the batch size (original was 32), these hyperparameters are kept the same as in the original paper. We also weight the model loss by the MF product, as was done with the original HyperSTARCOP models. As there is no validation subset, we use the model from the last epoch. All end-to-end models use 60 bands of AVIRIS-NG (namely, bands between 2104 and 2400 nm). This was chosen to match the range typically used by the matched filter as shown in Fig. 2.

Finally, when training models on the “OxHyperMinerals” dataset on the task of mineral detection, we train all models for 20 epochs with tiles of 64 px with 32 px overlap. In this case, trained models use all 285 bands of the available EMIT data (381–2492 nm). Here, the data loader is not balancing the samples and losses are not weighted.

A. Hardware Restrictions

For training, we use a high performance computing cluster with Tesla V100 GPUs with 32 GB GPU memory and the Intel Platinum 8628 (Cascade Lake), 2.90 GHz CPU with 384 GB RAM. This type of machine is typically much more computationally capable than the hardware later used for deployment.

Therefore, with several tested model variants, we also conducted experiments measuring the runtime on compute constrained hardware, namely, on the flatbed hardware provided by Unibap and D-Orbit. These machines serve as a proxy to match the computational in-space environment of a D-Orbit’s ION-SCV004 satellite, which is equipped with: quad-core 1.2 GHz, AMD GX-412HC SOC CPU processor, the Intel Movidius Myriad X VPU and 2 GB of RAM. The power consumption of the used environment (SpaceCloud iX5-100), is estimated to about 18 W by [39]. The same compute environment was used in [45] and similar environments leveraging Myriad chips were used on prior missions, such as ESA’s Phi-Sat-1 [53] and Phi-Sat-2 [54].

We have also selected two other devices typically used for Internet-of-Things applications, the CPU powered Raspberry Pi boards and the GPU equipped NVIDIA Jetson AGX Xavier board. As has been noted by NASA’s guidelines on using Raspberry Pi’s in space [55], they are not suggested for deployment in any mission critical tasks due to their lower robustness in extreme environments. They have been however used as proxies for devices of similar compute power in [56]. The estimated power consumption of Raspberry Pi devices is around 5 W

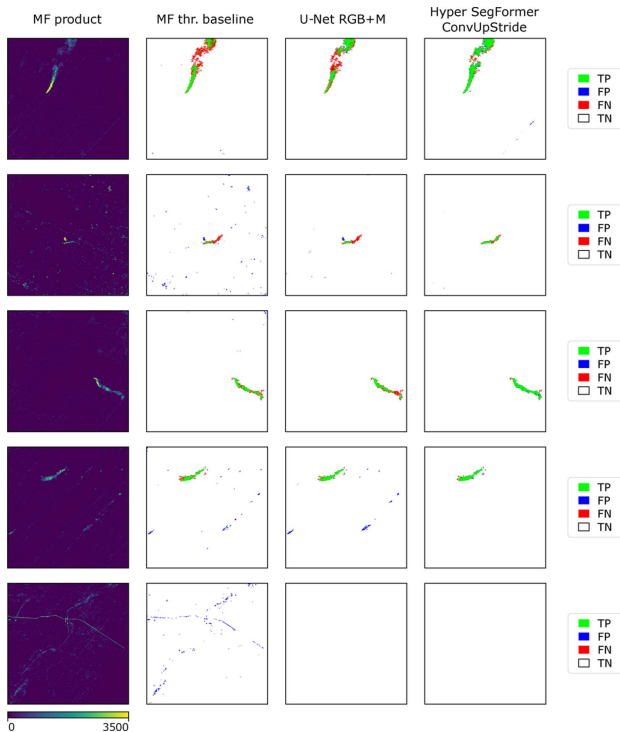


Fig. 6. Qualitative results on the OxHyperSyntheticCH4 dataset showing the computed matched filter in the first column and comparing the performance of the matched filter baseline, U-Net-based model from [17] and our proposed Hyper SegFormer model (“ConvUpStride” variant).

during a benchmark. We then use the more powerful NVIDIA Jetson AGX Xavier board equipped with an 8-core NVIDIA Carmel ARMv8.2 CPU at 2.265 GHz and a GV10B GPU based on the Volta architecture with 16 GB memory (shared between CPU and GPU). The power consumption of this board can be adjusted, in our case we used the 30 W mode. We used both ONNX Runtime python libraries and the Tensor RT inference engine (using the *trtexec* command), as was also used in the work of [2]. While initial tests of the effects of radiation on NVIDIA’s Jetson devices was explored in [57] and proposed for space applications in [58], these works also state that a careful examination and future tests are still needed.

V. RESULTS

In this section, we first compare the results from different used models on the task of methane detection in hyperspectral data from two explored sensors. Secondly, we report the results from using our proposed models on a novel dataset for mineral identification. Finally, we also report measurements of the inference time needed to use these models on a compute constrained hardware of the Unibap flatsats. These will serve as an approximate proxy to the runtime needed on-board of potential satellites.

A. Methane Detection on EMIT

Fig. 6 shows the comparison on the selected methods—classical baseline, U-Net-based model from [17] and our proposed Hyper SegFormer model (using the “1×1 Conv, Up,

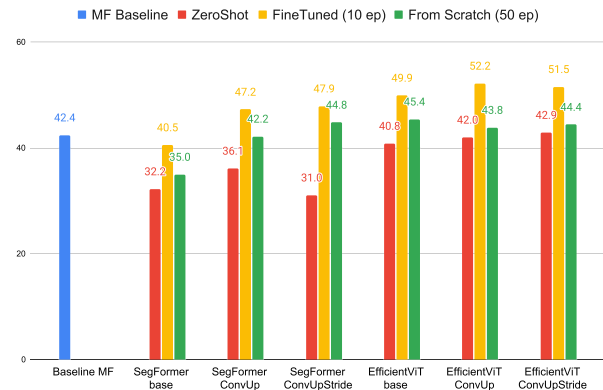


Fig. 7. Quantitative results on the OxHyperRealCH4 dataset showing the F1 score. We compare the MF baseline against zero-shot generalization, fine-tuning or training from scratch for both the proposed models. Results are averaged from 5 training runs, and the values on top of individual bars are rounded for better clarity.

Stride” variant). We illustrate the advantage of our proposed solution. The baseline and the U-Net-based models depend on the matched filter product (also shown in the first column), and as such, if this product does not capture the full extent of the methane plume present in the data, these models are not able to correctly reconstruct it. Meanwhile, our proposed model has access to the original source data and can extract information which would be lost when using the matched filter product. We show this in the first four rows. The last row demonstrates that if the matched filter product contains false detections (confounder materials such as roads, and building outlines), both learned models can filter these out.

Table II shows the results of several tested model variants for both SegFormer and EfficientViT architectures compared with classical machine learning approach of [17] and baseline method using [24] (regardless of the choice of kernel used in the baselines, although the “ones” baseline variant improves most of the scores over “cross” originally used in [17]). We note that all of the deep learning approaches significantly outperform the classical baseline. Furthermore, our proposed HyperSegFormer (ConvUpStride variant) gains over 43.85% in the IoU score (and 27.88% in F1 and 20.7% in AUPRC) over the HyperSTARCOP MF+RGB model. Our model also achieves over 155% in IoU score (and 96.9% in F1 score) over the classical matched filter baseline. These improvements hold over all measured metrics. We see increasing scores when implementing proposed adaptations in both the SegFormer (from 60.26 F1 score to 74.27) and EfficientViT (from 68.40 F1 score to 72.26) of our HyperspectralViT models.

In Fig. 7, we evaluate our adapted SegFormer and EfficientViT models on the OxHyperRealCH4 dataset of real methane leak events captured by the EMIT sensor. Fig. 8 shows the qualitative results on this dataset. Here, we use the models pretrained on the synthetic dataset (which uses data of the same input/output shapes and uses the same data normalization) and test their zero-shot generalization on the dataset of real events. Furthermore, we fine-tune these models on the smaller dataset of real events. Finally, we also compare these results with training models from

TABLE II
QUANTITATIVE RESULTS ON THE OXHYPERSYNTHETICCH4 DATASET

	AUPRC \uparrow	F1 \uparrow	Precision \uparrow	Recall \uparrow	IoU \uparrow	FPR by tile \downarrow
Baseline, mag1c, morpho. (cross 3x3)	N/A	37.65	31.69	46.37	23.19	89.0
Baseline, mag1c, morpho. (ones 3x3)	N/A	41.46	41.68	41.25	26.15	79.0
HyperSTARCOP MF+RGB [17]	68.10 ± 4.7	58.08 ± 5.3	48.44 ± 6.7	73.31 ± 2.8	41.12 ± 5.2	65.6 ± 6.7
SegFormer base	67.76 ± 2.6	60.26 ± 1.7	53.14 ± 4.9	70.70 ± 5.8	43.15 ± 1.7	27.2 ± 3.1
SegFormer ConvUp	78.14 ± 3.0	71.02 ± 3.4	68.72 ± 7.4	74.17 ± 3.2	55.18 ± 4.2	25.2 ± 7.2
SegFormer ConvUpStride	82.24 ± 2.7	74.27 ± 2.9	74.15 ± 4.0	74.88 ± 6.0	59.15 ± 3.7	18.4 ± 2.4
EfficientViT base	74.29 ± 7.3	68.40 ± 6.9	65.54 ± 10.3	72.75 ± 6.0	52.39 ± 7.8	25.0 ± 5.0
EfficientViT ConvUp	77.86 ± 3.4	70.89 ± 3.1	68.31 ± 5.8	74.06 ± 3.1	55.00 ± 3.7	40.2 ± 9.2
EfficientViT ConvUpStride	78.94 ± 5.3	72.26 ± 4.9	72.63 ± 6.3	72.19 ± 5.5	56.80 ± 6.0	31.8 ± 5.9

We show the average of training 5 runs of our models. We also compare the results against the HyperSTARCOP model from [17].

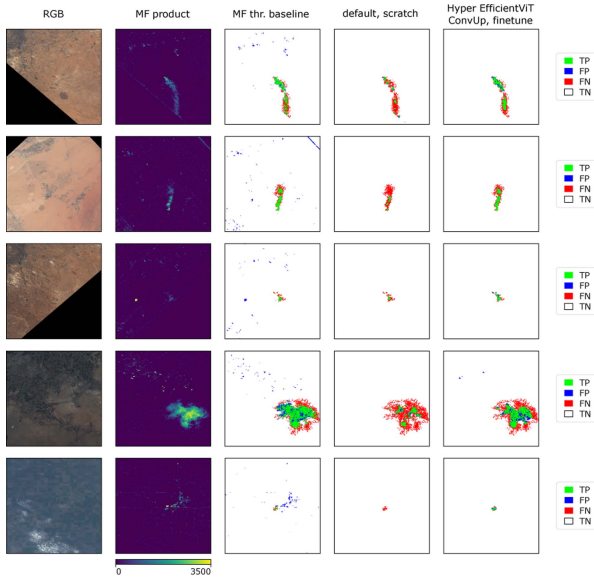


Fig. 8. Qualitative results on the OxHyperRealCH4 dataset. From left to right we show the predictions of the classical baselines, models trained from scratch and models pretrained on the synthetic dataset (using the HyperEfficientViT model, ConvUp variant). We note that black parts are the no-data areas created upon orthorectification.

scratch only on the dataset of real events. We note that this experiment resembles a real-world scenario, where we want to train models for a newly deployed sensor, which has not yet seen enough of these rare events. We try to answer the question of if pretraining on a synthetically created dataset helps the final models detect events on a real but small dataset.

Interestingly, when evaluating the models pretrained on the dataset with synthetic events on a new dataset of real methane leak events collected by EMIT, we observe that the zero-shot generalization performance is low (only two model variants reach the MF baseline—the “cross” kernel variant). We however see that with fine-tuning for ten epochs, we are able to significantly outperform both the classical baselines and all model variants trained from scratch. Namely, the HyperSegFormer (ConvUpStride variant) improves the performance of the MF baseline by 12.9% in F1 score (and 17.1% in IoU) and the same version of the model trained from scratch by over 6.9% in F1 score (and 8.8% in IoU). The HyperEfficientViT (ConvUp variant) outperforms the MF baselines by more than 23% in F1 (31.5% in

IoU) and the same model variant trained from scratch by 19.2% in F1 (and 25.6% in IoU). Interestingly, the ConvUp variant slightly outperforms the ConvUpStride variants on this dataset with the EfficientViT models. Overall, we see, that on the task of detecting real methane leak events in EMIT data, we benefit from pretraining our models on a dataset of synthetically created events (especially in cases similar to ours, where the dataset of real events is relatively small).

B. Methane Detection on AVIRIS-NG

In Table III, we evaluate our newly proposed end-to-end models on the benchmark dataset STARCOP [17] of real methane leak events in data from the aerial AVIRIS-NG. We train our proposed end-to-end models from scratch matching the settings used in the original work of [17].

When training our proposed models on the previously created benchmark dataset of real methane leak events in AVIRIS-NG data, we see some interesting mixed outcomes. While the performance of our models drops on F1 score of strong methane leak events (plume events larger than 1000 px) in contrast to the HyperSTARCOP “MF+RGB” model, on the metrics that aggregate all event sizes together, we see improvements. Namely, the HypeSegFormer (ConvUpStride variant) outperforms the HyperSTARCOP “MF+RGB” baseline by 12.9% in F1 score and by 18% in IoU score. The threshold-less AUPRC metric is also improved by over 17%. The high performance of previous STARCOP models on strong methane leak events could be explained by the fact that the used matched filter method, mag1c, works better for large plumes (as it is iteratively increasing the contrast of positive pixels—this could however lead to lowered performance for weaker events).

We have encountered very large gains on the synthetic and real EMIT datasets, however smaller ones on the benchmark dataset of AVIRIS-NG data. We note that perhaps better versions of the MF product could lead to stronger baselines also in the EMIT data. For example, Foote et al. [59] explored scene specific adaptation of MF products, and similarly there could be some EMIT specific improvements for calculating MF. This difference could also be caused by the number of used bands—while on the STARCOP dataset the end-to-end models use 60 bands, on the new EMIT datasets we use 86 bands, which also have a wider spectral range coverage.

TABLE III
QUANTITATIVE RESULTS ON THE BENCHMARK STARCOP DATASET

	AUPRC \uparrow	F1 \uparrow	F1 (strong) \uparrow	F1 (weak) \uparrow	Precision \uparrow	Recall \uparrow	IoU \uparrow	FPR by tile \downarrow
Baseline, mag1c, morpho. (cross 3x3)	N/A	40.14	67.50	39.95	30.57	58.42	25.11	75.43
Baseline, mag1c, morpho. (ones 3x3)	N/A	45.01	64.93	43.06	40.32	50.94	29.04	57.71
HyperSTARCOP MF+RGB [17]	51.99 \pm 2.8	50.26 \pm 3.8	81.96 \pm 3.7	43.42 \pm 5.7	37.53 \pm 5.8	78.39 \pm 6.6	33.63 \pm 3.4	43.66 \pm 7.4
SegFormer base	48.84 \pm 3.3	37.88 \pm 6.5	68.20 \pm 3.2	17.91 \pm 4.7	25.71 \pm 6.3	75.58 \pm 4.0	23.52 \pm 5.0	32.80 \pm 12.5
SegFormer ConvUp	57.91 \pm 5.6	53.17 \pm 6.5	71.53 \pm 4.0	39.37 \pm 8.7	44.93 \pm 10.8	68.16 \pm 5.9	36.42 \pm 6.0	32.34 \pm 20.2
SegFormer ConvUpStride	60.98 \pm 4.5	56.75 \pm 3.7	73.16 \pm 1.9	45.65 \pm 3.0	50.87 \pm 7.7	65.73 \pm 5.8	39.69 \pm 3.5	19.89 \pm 6.1

We show the average of training 5 runs of our models.

TABLE IV
QUANTITATIVE RESULTS ON THE MINERAL DATASET OXHYPERMINERALS

	F1	Precision	Recall	F1 Goethite	F1 Hematite	F1 Kaolinite
SegFormer base	81.77 \pm 1.54	77.33 \pm 3.58	86.92 \pm 1.79	79.31 \pm 1.08	80.88 \pm 0.66	85.79 \pm 3.84
SegFormer ConvUp	82.99 \pm 0.92	80.55 \pm 1.01	85.64 \pm 2.26	79.36 \pm 1.55	81.47 \pm 1.82	89.15 \pm 0.89
SegFormer ConvUpStride	84.70 \pm 0.70	83.42 \pm 2.69	86.14 \pm 1.98	81.19 \pm 1.03	83.59 \pm 1.41	90.35 \pm 0.80
EfficientViT base	77.08 \pm 0.74	79.37 \pm 3.73	75.28 \pm 3.99	69.88 \pm 3.09	77.89 \pm 1.09	85.25 \pm 1.05
EfficientViT ConvUp	79.73 \pm 0.92	82.66 \pm 2.32	77.11 \pm 2.39	74.59 \pm 1.28	79.74 \pm 0.67	86.55 \pm 1.30
EfficientViT ConvUpStride	79.18 \pm 1.16	84.03 \pm 2.74	74.96 \pm 2.17	73.57 \pm 1.20	77.93 \pm 2.70	87.24 \pm 1.22

We show the average of training 5 runs of our models.

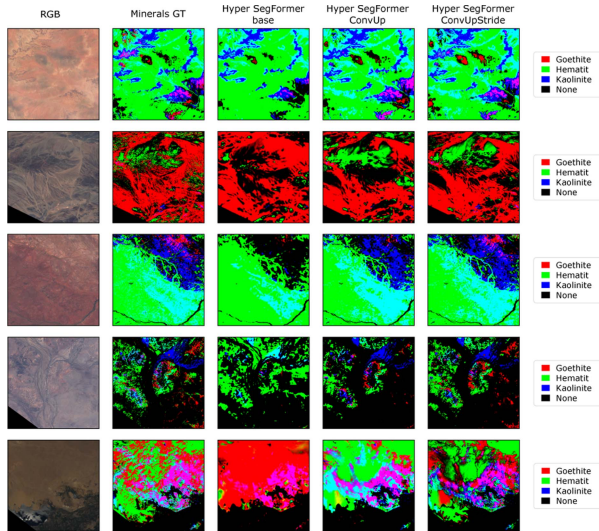


Fig. 9. Qualitative results comparing our new models on the mineral dataset OxHyperMinerals showing our proposed HyperSegFormer model variants (with output resolution degradation in default \times , “ConvUp” \times , and “ConvUpStride” full $1\times$).

C. Mineral Identification

In Table IV, we show the results of training the proposed machine learning architectures on the mineral dataset OxHyperMinerals. We show the individual F1 scores for each of the three mineral classes alongside the aggregated scores (which are support weighted by the number of pixels belonging to each class). In Fig. 9, we also show examples of model predictions next to the pseudoground truth generated by the Tetracorder system. Note the gradual increase of output resolution depending on the used adaptations.

When exploring the results of models trained on the dataset of minerals, we note smaller but sustained gains in performance due to the proposed adaptations. Namely, the HyperSegFormer

(ConvUpStride variant) improves the F1 score by over 3.5% in contrast to the vanilla model (and precision by 7.9%). The HyperEfficientViT demonstrates weaker performance for both base and adapted versions, however the used adaptations do improve this performance—the ConvUp variant scores 3.4% improvement in the F1 score.

Perhaps better improvement can be seen on the qualitative comparison between the models on Fig. 9, where we show the effect of gradually improving output resolution with the HyperSegFormer model. The model variants using “Up” and “Stride” adaptations (and their combination) show sharper predictions than the vanilla version of the model (as that one is limited to the output size reduced by the factor of 4).

We note that our experiments with mineral identification are initial, as we had to rely on pseudoground truth labels generated by Tetracorder.

D. Inference Times on Low Compute Hardware

Table V shows the model inference times needed to process a full EMIT scene capture of 1280×1242 px with our models for methane leak detection. We tile this full granule into 100 nonoverlapping tiles (padding the data if needed)—we note that in this way we could circumvent limitations of fixed input sizes encountered for example with the Myriad X VPU chip. These measurements show the IO time needed to load the hyperspectral datacube and the model inference time—for baselines this also includes the time needed to calculate the MF product.

We note that the matched filter computation is the slowest step influencing both the classical baselines and the HyperSTARCOP models. Our proposed end-to-end model variants avoid this step gaining vastly reduced inference time. Namely, our proposed HyperSegFormer ConvUp variant reduces the required time by over 85%, while the ConvUpStride variant reduces the time by over 41%—both compared against the HyperSTARCOP model and effectively also the classical MF baseline.

TABLE V
TIMING OF MODELS ON DIFFERENT DEVICES

	Params. (M)	Unibap CPU	Rasp3 CPU	Rasp4 CPU	Xavier CPU	X. GPU (onnx)	X. GPU (trtexec)
HyperSTARCOP MF+RGB [17]	6.633	203.3	1025.25	286.57	377	371.17	370.27
SegFormer Base	3.845	15.3	38.82	9.19	7.87	1.62	0.49
SegFormer UpConv	4.326	30.1	105.43	28.7	13.92	1.97	0.64
SegFormer UpConvStride	4.326	119.5	399.76	107.87	48.38	2.95	1.37
EfficientViT Base	4.811	N/A	39.96	9.45	9.39	2.36	0.37
EfficientViT ConvUp	4.855	N/A	47.73	13.81	9.99	2.00	0.55
EfficientViT ConvUpStride	4.855	N/A	91.07	21.26	18.09	2.10	0.5
IO	N/A	4	39.99	12.88	6.13	6.13	6.13

We estimate the time in seconds for processing a full granule of EMIT Data (1280×1242px tiled into 100 tiles of 128×128px). EfficientViT models were not evaluated on the Unibap machine. IO time to load the hyperspectral tiles is kept separate in the last row. We also illustrate the number of parameters each model uses.

When considering other devices, the MF calculation remains to be the bottleneck and the end-to-end models are faster. While using the CPU of the Raspberry Pi 3B+ the runtime is slower than on the Unibap flatbed, Raspberry Pi 4B reaches similar performance. Finally, the most powerful CPU of the NVIDIA’s Jetson AGX Xavier board shows the fastest runtimes of the tested CPU devices.

If we use our proposed HyperSegFormer ConvUp model on the Jetson AGX Xavier GPU, we achieve over 47× times speed-up over the Unibap’s machine CPU (e.g., 0.64 instead of 30.1 s per capture) and 317× times speed-up over the HyperSTARCOP model (203.3 s per capture). Using the GPU of the Jetson AGX Xavier board provides the fastest runtimes from all tested devices. We observed further speedup when using the Tensor RT engine (using the “trtexec” tool) and compiling the neural network directly on the device with allowed post-training mixed precision quantization (allowing float32 and float16)—in the case of the HyperSegFormer ConvUp, this is by over 70% in contrast to using the Onnxruntime library (without compilation and without mixed precision mode).

When using the Tensor RT inference engine, we also experience faster inference times when using the adapted EfficientViT architecture instead of the adapted SegFormer architecture—this is consistent with the fact the EfficientViT was optimized for fast inference speed in [2]. Namely, for the ConvUpStride variants, the HyperEfficientViT has runtime lower by about 64% in contrast with the HyperSegFormer models. Similar reduction can also be observed on the CPU of the Raspberry Pi devices.

For reference, the EMIT sensor collects up to 300 granules per day (and usually much less), and the NASA’s planned SBG mission [60] will produce an estimate of about 10x more. These granules have 1280×1242 px and can be tiled into 100 tiles of 128×128, which we used as reference when measuring model runtimes. With the estimated 300 granules and using the measurements from the Unibap machines, the mag1c matched filter baseline would run for almost 17 h (with 203 s for each capture), while the proposed ConvUp variant of our HyperSegFormer would take just 2.5 h (with 30 s for each capture). This time makes it more feasible to catch up with the estimated rate of data capture, we also assume that data loading could be handled on another thread in a delayed pipeline as in [61]. Further reduction to just estimated 3.2 min (with 0.64 s for each capture) could be achieved when using the GPU on the Jetson Xavier board.

This outlines the potential speed-up future space missions could leverage if using high-power compute units (even if these are turned on just for short periods of time).

Finally, we note that while the reported results measure inference times of the methane leak detection models, these findings also generalize to our mineral identification models. Slower inference times can be expected due to the increased spectral dimensionality of the used data (285 bands for the mineral identification models in contrast to 86 bands used by the models for methane detection).

VI. CONCLUSION

To summarize, our results indicate several interesting findings. First, we release HyperspectralViTs, the Transformer-based architectures adapted for the domain of hyperspectral data. The adapted models outperform the default versions and baselines both in accuracy scores and in inference speeds on-board of devices that serve as a proxy for the compute environment of the ION-SCV004 satellite.

Second, we demonstrate new state-of-the-art results in methane detection using our end-to-end machine learning architectures adapted for hyperspectral data. Namely, on a newly created dataset of synthetic events in data from the EMIT sensor, we improve the F1 score by more than 27% against the prior models depending on classical methane enhancement products from [17] and by over 97% in F1 score over classical matched filter baselines. When comparing with off-the-shelf performance, our proposed HyperspectralViT variant of the SegFormer model improves the F1 score by over 23% on this synthetic dataset. We confirm our findings also on a benchmark dataset released by [17], where we improve the overall F1 score by almost 13% and the AUPRC score by more than 17%. Finally, on a newly created dataset of real methane leak events recorded by the EMIT sensor, we demonstrate, that it is possible to leverage models pretrained on synthetic data by fine-tuning them on smaller real datasets. Using the best performing model variant, we show that the finetuned model improves the F1 score by 6.9% over models trained from scratch.

On a newly created dataset for the task of mineral identification, we note that our HyperSegFormer (ConvUpStride variant) improves the F1 score by over 3.5% in contrast to the vanilla model. More notably, we show the influence of the retained

output resolution. While the default versions of the Transformer based models lose details in predicted scenes, our adaptations greatly improve the quality of the predictions. We note that our results are conducted on an initial dataset with pseudoground truth labels made by the Tetracorder method [27]. However, with a manually verified dataset, it will be easy to retrain our models. Our initial results pave the path for future experiments - to support this, we release our dataset with labels for all separate constituents.

Furthermore, our models could be leveraged as Foundation models initially trained on the detection of individual chemical components and later finetuned or adapted for a variety of downstream tasks (methane leak detection included). Gained performance on the general detection task could also help the other tasks of interest, as was exemplified in the case of multitask learning [62].

In terms of inference speed, we achieve a 85.19% faster performing model against the previously reported models and baselines of [17]. We report the inference times measured on hardware, which is a realistic proxy to the currently flying ION-SCV 004 satellite. We highlight that with our model, one scene can be processed in around half a minute ($6.7\times$ times speedup against the baseline). We also measure our models on several commercial off-the-shelf devices, these exhibit further potential speedups and have started to be explored for space missions in [63]. Namely, we show over $317\times$ times speedup when using GPUs on the NVIDIA Jetson AGX Xavier board.

As a potential limitation of the proposed end-to-end HyperspectralViT models, they operate directly on the bands of hyperspectral data and as such, they cannot be used for zero-shot generalization as for example the model of [17]. We however note, that a similar simulated dataset can be created from representative data for any other satellite and later used to train end-to-end models. Second, our model does not support additional tasks, such as plume quantification or source point estimation. We however note, that the prediction of our model can be used as a binary mask to clean the noisy methane enhancement products, after which classical pipelines for these tasks can be used [64].

As a further limitation, we also note, that we used the L1B level of data processing in our experiments instead of the truly “raw” versions of the data (sometimes called L0). This is due to the fact that this data is not available for these sensors, in fact this is a common problem in the wider research of methods for on-board data processing. Only recently, small datasets for multispectral sensors were made available in the works of [65]. However, works, such as [66], show that real-time correction and preprocessing of the L0 hyperspectral data are feasible on low-cost and low-compute platforms. Similarly, Castano et al. [67] used partially processed data (so-called level 0.5) on-board of the EO-1 spacecraft. Furthermore, works, such as [68], demonstrate that tiny machine learning models can serve as simulators for fast translation between processing levels (in that case L1C to L2A of Sentinel-2 data). For real world deployment, we’d suggest two approaches: either using even partial data processing pipelines on-board, or training deep models with simulated L0-like data.

For future research, we would like to explore further steps to make the models more efficient. Our proposed changes to the

Transformer-based models model can be combined with pruning reported in [40] and [69]. Specialized accelerators as those suggested in [70] could also be leveraged. Secondly, we would also like to explore extension of our models with methods for uncertainty estimation. Methods, such as model ensembles [71], or Monte Carlo Dropout [72], could be easily added to the proposed semantic segmentation models, as was previously demonstrated in [73]. Third, we note that the existing archives of EMIT data (currently with more than 121 k captures totaling more than 244 TB of data) offer unique opportunities to train large general foundation models on hyperspectral data, where one of the example tested tasks could also be methane detection. Models and datasets proposed in this article could likely be leveraged for general learning tasks (in the direction outlined by our mineral identification task), but also for other downstream applications highlighted by the future planned missions, such as NASA’s SBG [60] or ESA’s CHIME [74].

Finally, we are releasing the three newly created datasets¹ split into several repositories—OxHyperRealCH4, OxHyperSyntheticCH4, and OxHyperMinerals.

ACKNOWLEDGMENT

The authors would like to thank D-Orbit and Unibap for access to the SpaceCloud Hardware when measuring our models inference speeds on realistic hardware proxy of a real satellite environment.

REFERENCES

- [1] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, “SegFormer: Simple and efficient design for semantic segmentation with transformers,” in *Proc. 35th Int. Conf. Neural Inf. Process. Syst.*, 2021, pp. 12077–12090.
- [2] H. Cai, J. Li, M. Hu, C. Gan, and S. Han, “EfficientViT: Lightweight multi-scale attention for high-resolution dense prediction,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 17302–17313.
- [3] J. C. Kuylensstierna, E. Michalopoulou, and C. Malley, “Global methane assessment: Benefits and costs of mitigating methane emissions,” 2021. [Online]. Available: <https://coilink.org/20.500.12592/00gj20>
- [4] N. A. A. Braham, C. M. Albrecht, J. Mairal, J. Chanussot, Y. Wang, and X. X. Zhu, “SpectralEarth: Training hyperspectral foundation models at scale,” 2024, *arXiv:2408.08447*.
- [5] D. Wang et al., “Hypersigma: Hyperspectral intelligence comprehension foundation model,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2025, pp. 1–18.
- [6] S. Kumar, I. Arevalo, A. Iftexhar, and B. Manjunath, “Methanemapper: Spectral absorption aware hyperspectral transformer for methane detection,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 17609–17618.
- [7] G. Furano et al., “Towards the use of artificial intelligence on the edge in space systems: Challenges and opportunities,” *IEEE Aerosp. Electron. Syst. Mag.*, vol. 35, no. 12, pp. 44–56, Dec. 2020.
- [8] J. Parr et al., “Live twinning: A vision of ML enabled assets in leo for rapid response to natural catastrophes,” in *Proc. IGARSS 2024-2024 IEEE Int. Geosci. Remote Sens. Symp.*, 2024, pp. 685–688.
- [9] G. Mateo-Garcia et al., “Towards global flood mapping onboard low cost satellites with machine learning,” *Sci. Rep.*, vol. 11, no. 1, 2021, Art. no. 7249.
- [10] S. Peyghambari and Y. Zhang, “Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: An updated review,” *J. Appl. Remote Sens.*, vol. 15, no. 3, 2021, Art. no. 031501.

¹[Online]. Available: <https://huggingface.co/previtus>

- [11] B. Lu, P. D. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sens.*, vol. 12, no. 16, 2020, Art. no. 2659.
- [12] D. L. Blaney et al., "The Europa Clipper mapping imaging spectrometer for Europa (MISE): Using compositional mapping to understand Europa," in *Proc. 50th Annu. Lunar Planet. Sci. Conf.*, 2019, Art. no. 2218.
- [13] B. L. Ehlmann et al., "NASA's Lunar Trailblazer Mission: A Pioneering Small Satellite for Lunar Water and Lunar Geology," in *Proc. IEEE Aerosp. Conf.*, 2022, pp. 1–14.
- [14] R. Green et al., "The moon mineralogy mapper (M3) imaging spectrometer for lunar science: Instrument description, calibration, on-orbit measurements, science data calibration and on-orbit validation," *J. Geophysical Res.: Planets*, vol. 116, no. E10, pp. 31, 2011.
- [15] A. P. Ravikumar, J. Wang, and A. R. Brandt, "Are optical gas imaging technologies effective for methane leak detection?," *Environ. Sci. Technol.*, vol. 51, no. 1, pp. 718–724, 2017.
- [16] Spectracity, "See the world in true colors: The first smartphone with multispectral imaging camera makes its global debut at CES," Accessed: Sep. 29, 2024. [Online]. Available: <https://spectracity.com/>
- [17] V. Růžička, G. M.-García, L. G.-Chova, A. Vaughan, L. Guanter, and A. Markham, "Semantic segmentation of methane plumes with hyperspectral machine learning models," *Sci. Rep.*, vol. 13, no. 1, 2023, Art. no. 19999.
- [18] R. Green, "EMIT L1B at-sensor calibrated radiance and geolocation data 60 m V001," Distributed by NASA EOSDIS Land Processes Distributed Active Archive Center, 2022, Accessed Sep. 29, 2024. [Online]. Available: <https://doi.org/10.5067/EMIT/EMITL1BRAD.001>
- [19] E. Tiemann, S. Zhou, A. Kläser, K. Heidler, R. Schneider, and X. X. Zhu, "Machine learning for methane detection and quantification from space - a survey," 2024. [Online]. Available: <https://arxiv.org/abs/2408.15122>
- [20] L. Brown et al., "Methane line parameters in the HITRAN2012 database," *J. Quantitative Spectrosc. Radiative Transfer*, vol. 130, pp. 201–219, 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022407313002744>
- [21] D. J. Varon, D. Jervis, J. McKeever, I. Spence, D. Gains, and D. J. Jacob, "High-frequency monitoring of anomalous methane point sources with multispectral Sentinel-2 satellite observations," *Atmospheric Meas. Techn.*, vol. 14, no. 4, pp. 2771–2785, Apr. 2021. [Online]. Available: <https://amt.copernicus.org/articles/14/2771/2021/>
- [22] E. Sánchez-García, J. Gorroño, I. I.-Loitxate, D. J. Varon, and L. Guanter, "Mapping methane plumes at very high spatial resolution with the WorldView-3 satellite," *Atmos. Meas. Techn.*, vol. 15, no. 6, pp. 1657–1674, Mar. 2022. [Online]. Available: <https://amt.copernicus.org/articles/15/1657/2022/>
- [23] D. R. Thompson et al., "Real-time remote detection and measurement for airborne imaging spectroscopy: A case study with methane," *Atmos. Meas. Techn.*, vol. 8, no. 10, pp. 4383–4397, 2015. [Online]. Available: <https://amt.copernicus.org/articles/8/4383/2015/>
- [24] M. D. Foote et al., "Fast and accurate retrieval of methane concentration from imaging spectrometer data using sparsity prior," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6480–6492, Sep. 2020.
- [25] J. Roger, L. Guanter, J. Gorroño, and I. I.-Loitxate, "Exploiting the entire near-infrared spectral range to improve the detection of methane plumes with high-resolution imaging spectrometers," *Atmos. Meas. Techn. Discuss.*, vol. 17, pp. 1333–1346, 2024.
- [26] H. Shirmard, E. Farahbakhsh, R. D. Müller, and R. Chandra, "A review of machine learning in processing remote sensing data for mineral exploration," *Remote Sens. Environ.*, vol. 268, 2022, Art. no. 112750.
- [27] R. N. Clark et al., "Imaging spectroscopy: Earth and planetary remote sensing with the USGS tetracorder and expert systems," *J. Geophysical Res.: Planets*, vol. 108, no. E12, pp. 5131, 2003.
- [28] R. N. Clark, "spectroscopy of rocks and minerals, and principles of spectroscopy," in *Manual of Remote Sensing*. New York, NY, USA: Wiley, 1999, pp. 3–58.
- [29] Y.-q. Wan, Y.-h. Fan, and M.-s. Jin, "Application of hyperspectral remote sensing for supplementary investigation of polymetallic deposits in huanishan ore region, northwestern China," *Sci. Rep.*, vol. 11, no. 1, 2021, Art. no. 440.
- [30] H. Sun, X. Zheng, X. Lu, and S. Wu, "Spectral-spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3232–3245, May 2020.
- [31] D. Hong et al., "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5518615. [Online]. Available: <https://ieeexplore.ieee.org/document/9627165>
- [32] D. R. Thompson and P. G. Brodrick, "Realizing machine learning's promise in geoscience remote sensing," *EOS*, vol. 102, Jul. 2021. [Online]. Available: <https://doi.org/10.1029/2021EO160605>
- [33] A. Signoroni, M. Savardi, A. Baronio, and S. Benini, "Deep learning meets hyperspectral image analysis: A multidisciplinary review," *J. Imag.*, vol. 5, no. 5, 2019, Art. no. 52.
- [34] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 158, pp. 279–317, 2019.
- [35] U. B. Gewali, S. T. Monteiro, and E. Saber, "Machine learning based hyperspectral image analysis: A survey," 2018, *arXiv:1802.08701*.
- [36] M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe, "220 band AVIRIS hyperspectral image data set: Jun. 12, 1992 Indian pine test site 3," Sep. 2015. [Online]. Available: <https://purr.purdue.edu/publications/1947/1>
- [37] P. Joyce et al., "Using a deep neural network to detect methane point sources and quantify emissions from PRISMA hyperspectral satellite images," *Atmospheric Meas. Techn.*, vol. 16, pp. 2627–2640, May 2023.
- [38] I. Jahan, M. Mehana, G. Matheou, and H. Viswanathan, "Deep learning-based quantifications of methane emissions with field applications," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 132, 2024, Art. no. 104018.
- [39] O. Flordal et al., "Spacecloud cloud computing and in-orbit demonstration," in *Proc. Eur. Workshop On-Board Data Process.*, 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.5638576>
- [40] C. Yang et al., "Pruning parameterization with bi-level optimization for efficient semantic segmentation on the edge," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 15402–15412.
- [41] G. Giuffrida et al., "CloudScout: A deep neural network for on-board cloud detection on hyperspectral images," *Remote Sens.*, vol. 12, no. 14, 2020, Art. no. 2205.
- [42] S. Chien et al., "Using autonomy flight software to improve science return on Earth observing one," *J. Aerosp. Comput., Inf., Commun.*, vol. 2, no. 4, pp. 196–216, 2005.
- [43] D. R. Thompson et al., "Autonomous spectral discovery and mapping onboard the EO-1 spacecraft," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3567–3579, Jun. 2013.
- [44] V. Růžička et al., "Ravæn: Unsupervised change detection of extreme events using ML on-board satellites," *Sci. Reports*, vol. 12, no. 1, 2022, Art. no. 16939.
- [45] V. Růžička et al., "Fast model inference and training on-board of satellites," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 2002–2005, doi: [10.1109/IGARSS52108.2023.10282715](https://doi.org/10.1109/IGARSS52108.2023.10282715).
- [46] S. Jongaramrungruang, A. K. Thorpe, G. Matheou, and C. Frankenberg, "MethaNet—An AI-driven approach to quantifying methane point-source emission from high-resolution 2-D plume imagery," *Remote Sens. Environ.*, vol. 269, Feb. 2022, Art. no. 112809. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425721005290>
- [47] G. Matheou and D. Chung, "Large-Eddy simulation of stratified turbulence. Part II: Application of the stretched-vortex model to the atmospheric boundary layer," *J. Atmospheric Sci.*, vol. 71, no. 12, pp. 4439–4460, 2014.
- [48] R. N. Clark, G. A. Swayze, V. J. Realmuto, P. G. Brodrick, and D. R. Thompson, "EMIT L2b algorithm: Mineral detection and related products at the pixel scale [ATBD V0.2]," Distributed by NASA, Accessed Apr. 4, 2020.
- [49] R. Green, "EMIT L2B estimated mineral identification and band depth and uncertainty 60 m V001," Distributed by NASA EOSDIS Land Processes Distributed Active Archive Center, 2023, Accessed Aug. 16, 2024. [Online]. Available: <https://doi.org/10.5067/EMIT/EMITL2BMIN.001>
- [50] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Intervention*, 2015, pp. 234–241.
- [51] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. JPL, Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 147–149.
- [52] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.
- [53] G. Giuffrida et al., "The ϕ -sat-1 mission: The first on-board deep neural network demonstrator for satellite Earth observation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5517414.
- [54] A. Marin, C. Coelho, F. Deconinck, I. Babkina, N. Longepe, and M. Pastena, "Phi-Sat-2: Onboard AI apps for earth observation," in *Proc. Space Artif. Intell., Online Conf.*, Sep. 13, 2021.

- [55] S. M. Guertin, "Raspberry Pis for space guideline," NASA: Washington, DC, USA, 2022. [Online]. Available: <https://nepp.nasa.gov/docs/papers/2021-Guertin-Raspberry-Pi-Guideline-CL-21-5641.pdf>
- [56] E. W. Gretok and A. D. George, "Onboard multi-scale tile classification for satellites and other spacecraft," in *Proc. IEEE Space Comput. Conf.* 2021, pp. 110–121.
- [57] W. S. Slater, N. P. Tiwari, T. M. Lovelly, and J. K. Mee, "Total ionizing dose radiation testing of NVIDIA Jetson nano GPUs," in *Proc. IEEE High Perform. Extreme Comput. Conf.*, 2020, pp. 1–3.
- [58] M. Lofqvist and J. Cano, "Accelerating deep learning applications in space," in *Proc. 34th Annu. Small Satell. Conf., Virtual Conf.*, 2020.
- [59] M. D. Foote et al., "Impact of scene-specific enhancement spectra on matched filter greenhouse gas retrievals from imaging spectroscopy," *Remote Sens. Environ.*, vol. 264, 2021, Art. no. 112574.
- [60] K. Cawse-Nicholson et al., "NASA's surface biology and geology designated observable: A perspective on surface imaging algorithms," *Remote Sens. Environ.*, vol. 257, 2021, Art. no. 112349.
- [61] V. Růžička and F. Franchetti, "Fast and accurate object detection in high resolution 4K and 8K video using GPUs," in *Proc. IEEE High Perform. Extreme Comput. Conf.*, 2018, pp. 1–7.
- [62] I. Kokkinos, "UberNet: Training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory," in *Proc. IEEE Conf. Computer Vis. Pattern Recognit.*, 2017, pp. 6129–6138.
- [63] J. Swope et al., "Benchmarking remote sensing image processing and analysis on the snapdragon processor onboard the international space station," in *Proc. IGARSS IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 5305–5308.
- [64] A. K. Thorpe et al., "Mapping methane concentrations from a controlled release experiment using the next generation airborne visible/infrared imaging spectrometer (AVIRIS-NG)," *Remote Sens. Environ.*, vol. 179, pp. 104–115, 2016.
- [65] G. Meoni et al., "Unlocking the use of raw multispectral earth observation imagery for onboard artificial intelligence," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 12521–12537, 2024.
- [66] M. B. Henriksen, J. L. Garrett, E. F. Prentice, A. Stahl, T. A. Johansen, and F. Sigernes, "Real-time corrections for a low-cost hyperspectral instrument," in *Proc. 10th Workshop Hyperspectral Imag. Signal Processing: Evol. Remote Sens.*, 2019, pp. 1–5.
- [67] R. Castano et al., "Onboard classifiers for science event detection on a remote sensing spacecraft," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2006, pp. 845–851.
- [68] C. Aybar et al., "Onboard cloud detection and atmospheric correction with efficient deep learning models," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 19518–19529, 2024.
- [69] H. Bai, H. Mao, and D. Nair, "Dynamically pruning segformer for efficient semantic segmentation," in *Proc. ICASSP IEEE Int. Conf. Acoust., Speech Signal Process.*, 2022, pp. 3298–3302.
- [70] Y. Yousfia and D. Wischerta, "Spiking neural network design for on-board detection of methane emissions through neuromorphic computing," in *Proc. 75th Int. Astronautical Congr.*, 2024, pp. 1212–1223.
- [71] W. H. Beluch, T. Genewein, A. Nürnberger, and J. M. Köhler, "The power of ensembles for active learning in image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 9368–9377.
- [72] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.
- [73] V. Růžička, S. D'Aronco, J. D. Wegner, and K. Schindler, "Deep active learning in remote sensing for data efficient change detection," in *Proc. ECML/PKDD Workshop Mach. Learn. Earth Observ.*, 2020.
- [74] M. Rast, J. Nieke, J. Adams, C. Isola, and F. Gascon, "Copernicus hyperspectral imaging mission for the environment (CHIME)," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. IGARSS*, 2021, pp. 108–111.



Vít Růžička received the B.Sc. and M.Sc. degrees in computer science from the Czech Technical University, Prague, Czechia, in 2014 and 2017, respectively. In April 2025, he received his Ph.D. degree in computer science at the University of Oxford, Oxford, U.K.

He was a Visiting Researcher with the European Space Agency (2023) and a Consultant for the United Nations Environment Programme (2024). His research interests include machine learning applied to remote sensing applications and tasks connected to climate change and on-board deployment on low compute systems, such as satellites.



Andrew Markham received the B.Sc. (Hons) and Ph.D. degrees in electrical engineering from the University of Cape Town, Cape Town, South Africa, in 2004 and 2008, respectively.

He is a Professor of computer science with the University of Oxford, U.K. He has authored or coauthored more than 200 papers and works in the broad area of cyberphysical system. His research focuses on the application of sensing, systems, and signal processing to increased sustainability, conservation, and autonomy.