EoS-FM: Can an Ensemble of Specialist Models act as a Generalist Feature Extractor ?

Pierre Adorni[1], Minh-Tan Pham[1], Stéphane May[2], Sébastien Lefèvre[1,3]

1. *IRISA, Université Bretagne Sud, UMR 6074, Vannes, France*
2. *Centre National d'Études Spatiales (CNES), Toulouse, France*
3. *UiT The Arctic University of Norway, Tromsø, Norway*

The Problem

Recent advances in foundation models have shown great promise in the Remote Sensing field, **improving the SOTA** time and time again on numerous downstream tasks. However, this improvement is often **at the cost of power efficiency**, as the main method to increase performance has been **upscaling models and pretraining datasets** alike [1-2]. This rush for the biggest models poses multiple problems, such as rendering difficult the adaptation and usage of models in a low resources setting.

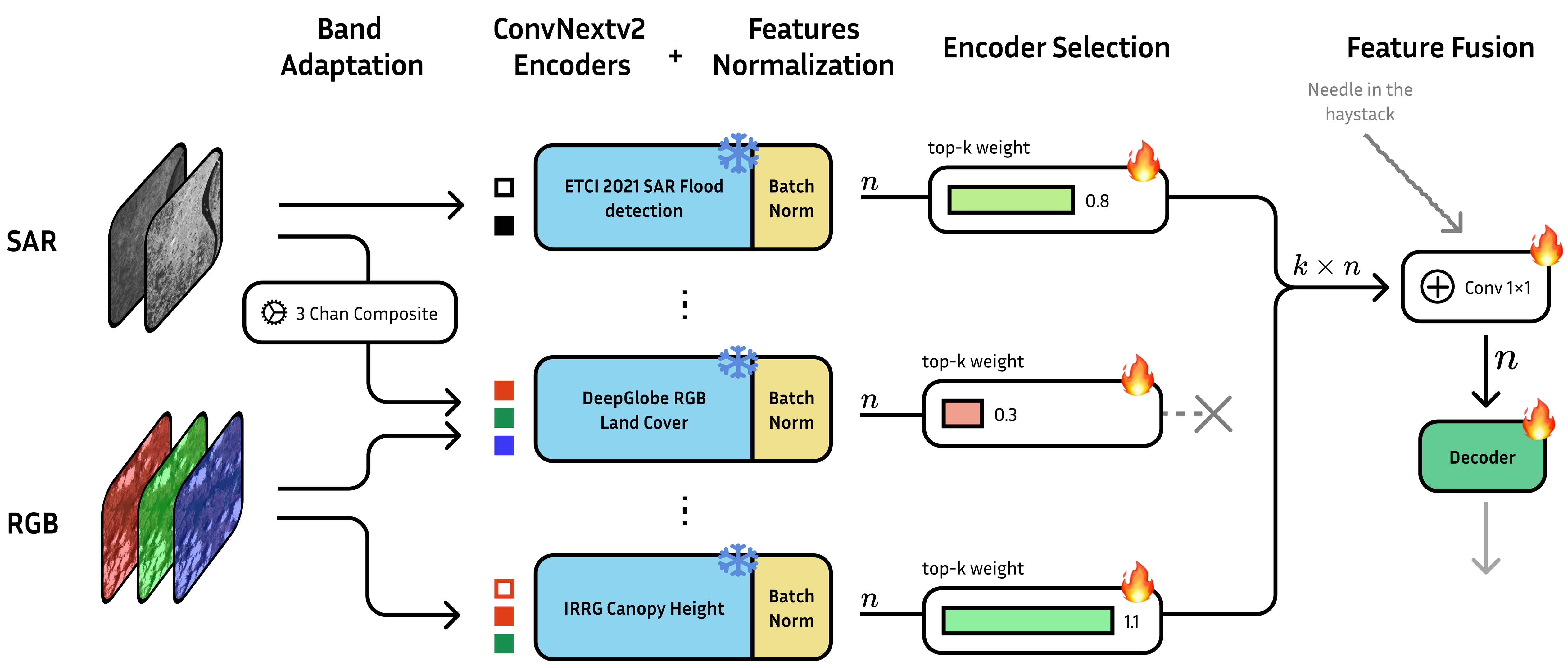

Figure: *Our EoS-FM Backbone adapts any given input to a multitude of formats using band duplication and selection to extract as many feature maps as possible, and then fuse them. Each encoder produces n feature maps; a subset of k encoders is then selected for fusion, and their $k \times n$ feature maps are aggregated into n fused feature maps before being passed to the decoder.*

Model	HLS Burns	MADOS	PASTIS	Sen1FL	xView2	FBP	DynEarthNet	CropMap	SN7	AI4Farms	BioMass ↓	Mean DTB ↓
Scale-MAE (303M) ❄️	75.47	21.47	22.86	64.74	**56.06**	48.75	35.27	13.44	49.68	26.66	54.16	13.09
GFM-Swin (87M) ❄️	67.23	28.19	21.47	62.57	53.45	55.58	28.16	27.21	39.48	32.88	49.30	12.48
SatlasNet (87M) ❄️	74.79	29.87	16.76	83.92	44.07	37.86	34.64	29.08	49.78	13.91	44.38	12.17
Prithvi (87M) ❄️	<u>77.73</u>	21.24	**33.56**	86.28	35.08	29.98	32.28	27.71	36.78	35.04	41.19	11.79
DOFA (112M) ❄️	71.98	23.77	27.68	82.84	<u>55.60</u>	27.82	**39.15**	<u>29.91</u>	46.10	27.74	46.03	10.70
CROMA (303M) ❄️	76.44	32.44	<u>32.80</u>	<u>87.22</u>	46.54	37.39	<u>36.08</u>	**36.77**	42.15	38.48	**40.25**	7.11
EoS-FM (72 M) ❄️	71.82	**47.05**	29.24	79.48	55.27	**64.18**	32.05	22.97	**52.13**	**40.20**	41.82	**4.70**
EoS-FM Small (22M) ❄️	71.48	<u>45.53</u>	26.41	80.16	54.06	<u>62.80</u>	32.59	21.83	<u>51.82</u>	<u>39.88</u>	47.11	<u>5.89</u>
UNet (~8M) 🔥	**79.46**	24.30	29.53	**88.55**	46.77	52.58	35.59	13.88	46.08	34.84	<u>40.39</u>	8.46

Table: *Results on the Pangaea benchmark [4], **using 10% of the training labels**. Our method has the best performance on four different datasets, and exhibits the best average performance, even surpassing supervised baselines (DTB = Distance To Best [5]).*
↓ means lower is better.

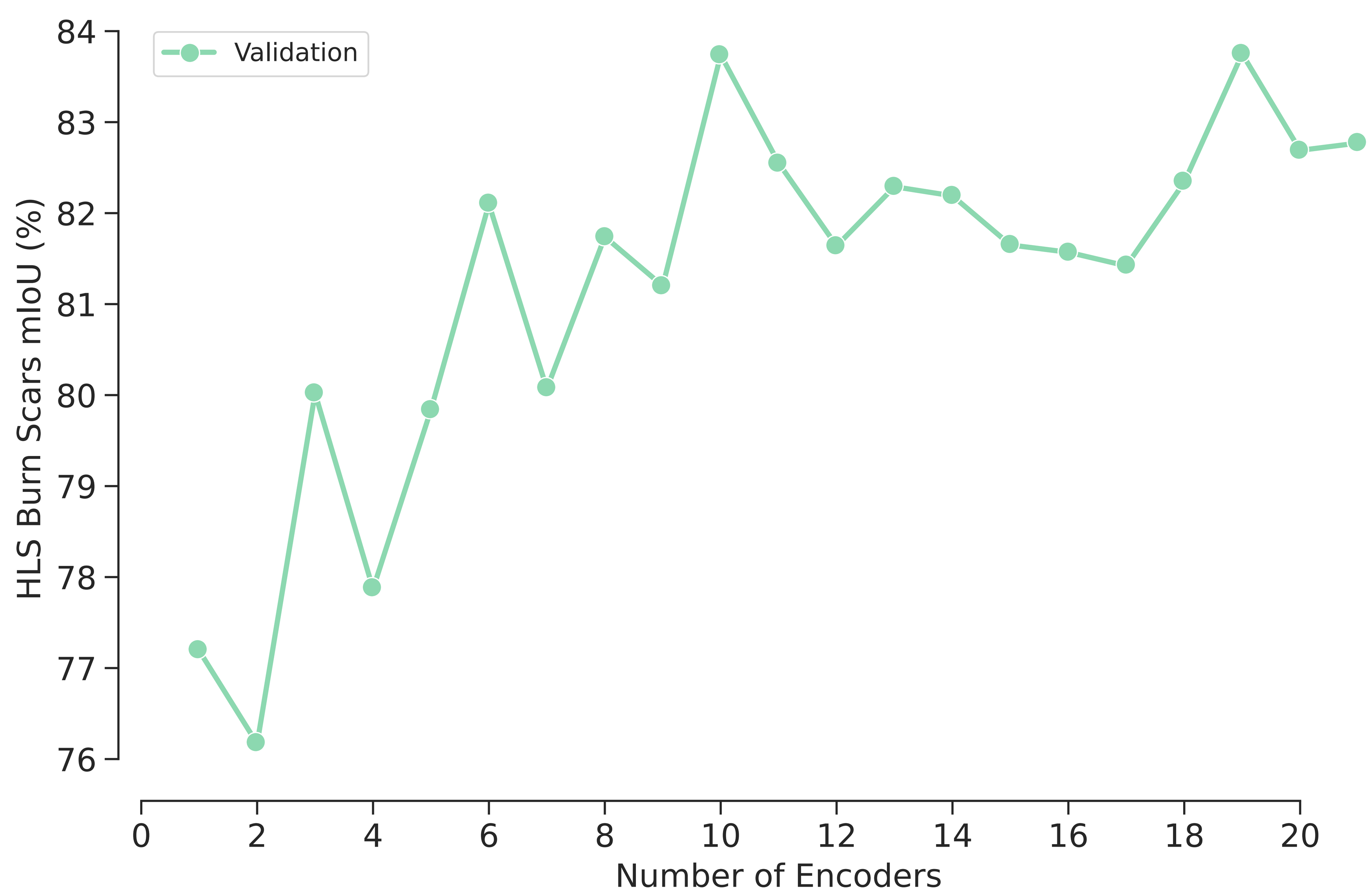

Figure: *Increasing the number of encoders increases the performance of the ensemble in a frozen setting.*

Key Takeaways

- An Ensemble of Specialists turns out to be a **realistic alternative** to large monolithic models, offering strong generalization capabilities, even in a regime of data scarcity.
- Due to its design, our proposed architecture can easily be **pruned** during training, and can easily be **improved** by adding more specialists. Furthermore, our architecture naturally supports **federated learning** setups.
- Together, these elements illustrate a different path toward **sustainable** and collaborative RSFM development, grounded in **flexibility, efficiency, and extensibility**.

[1] Keumgang Cha et al. "A Billion-scale Foundation Model for Remote Sensing Images." In JSTARS (2024), pp. 1–17. ISSN: 1939-1404, 2151-1535. DOI: 10 . 1109 / JSTARS . 2024 . 3401772. arXiv: 2304 . 05215 [cs].(Visited on 12/16/2024).

[4] V. Marsocci et al. "PANGAEA: A Global and Inclusive Benchmark for Geospatial Foundation Models", (arXiv:2412.04204), 2024.

[2] Philipe Dias et al. OReole-FM: Successes and Challenges toward Billion-Parameter Foundation Models for High- Resolution Satellite Imagery. In SIGSPATIAL. Oct. 2024. DOI: 10.48550/ arXiv . 2410 . 19965. arXiv: 2410 . 19965. (Visited on 11/14/2024).

[3] Shazeer Noam et al. "Outrageously large neural networks: The sparsely-gated mixture-of-experts layer." In ICLR (2017).

[5] Adorni Pierre et al. "Towards Efficient Benchmarking of Foundation Models in Remote Sensing: A Capabilities Encoding Approach." In CVPRW, pp. 3096-3106. 2025.

pierre.adorni@irisa.fr