Shaping Fine-Tuning of Geospatial Foundation Models: Effects of Label Availability and Temporal Resolution

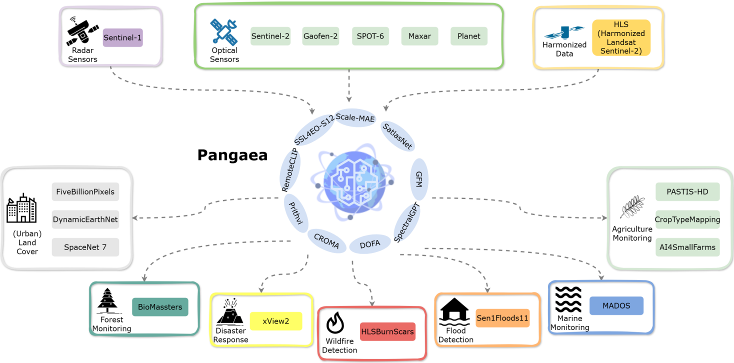
Giovanni Castiglioni, Nicolás Isla, Cristian B. Calderon, Javiera Castillo-Navarro, Sébastien Lefèvre, Valentin Barriere

ICML TerraBytes Workshop

Universidad de Chile, CENIA, Conservatoire National des Arts et Métiers, Université Bretagne Sud

Motivation

- Remote sensing is shifting from task-specific models to GFMs.

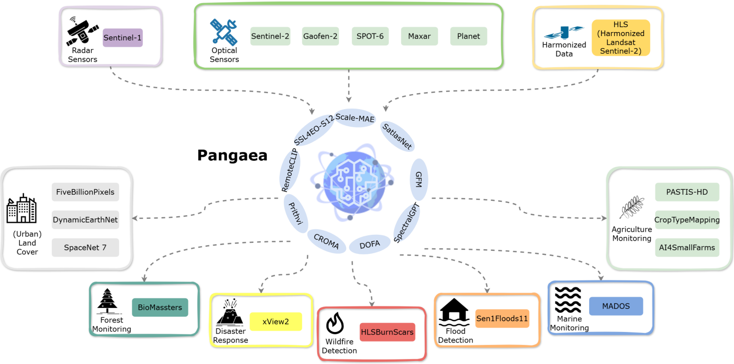


Evaluation on Diverse Downstream Datasets and Applications

Marsocci, V., et al. "PANGAEA: A Global and Inclusive Benchmark for Geospatial Foundation Models." 2024.

Motivation

- Remote sensing is shifting from task-specific models to GFMs.
- Scarcity of labeled data and temporal sequences is a challenge.

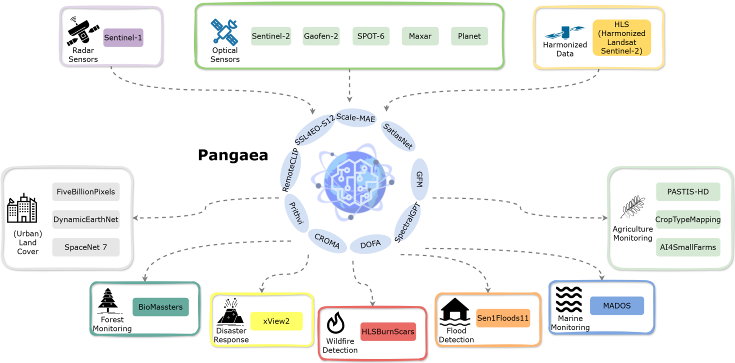


Evaluation on Diverse Downstream Datasets and Applications

Marsocci, V., et al. "PANGAEA: A Global and Inclusive Benchmark for Geospatial Foundation Models." 2024.

Motivation

- Remote sensing is shifting from task-specific models to GFMs.
- Scarcity of labeled data and temporal sequences is a challenge.
- **How should we fine-tune GFMs for a specific task?**

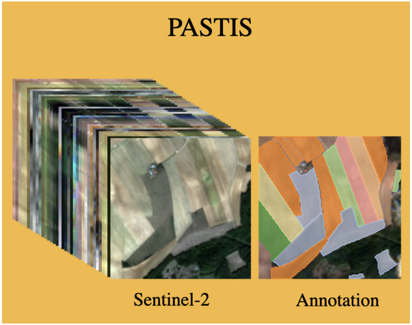


Evaluation on Diverse Downstream Datasets and Applications

Marsocci, V., et al. "PANGAEA: A Global and Inclusive Benchmark for Geospatial Foundation Models." 2024.

Dataset: PASTIS-HD

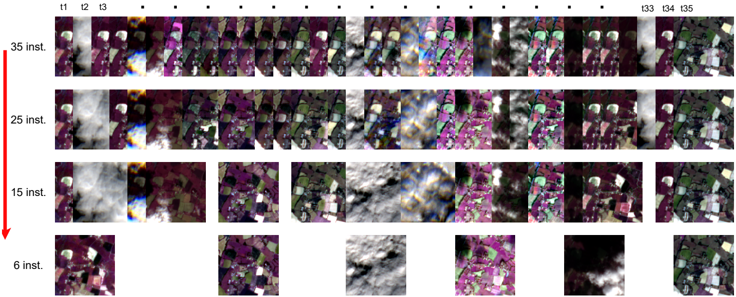
- **Crop-type mapping** of ∼2.4k SITS of 128x128 and t $\in$ [38,63].
- We used **Sentinel-2 images**, excluding S1 and SPOT-6 sensors.



Garnot, V.S.F., et al. "Multi-Modal Temporal Attention Models for Crop Mapping from Satellite Time Series." ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 187, pp. 294-305, 2022.

Dataset: PASTIS-HD

- **Crop-type mapping** of ∼2.4k SITS of 128x128 and t ∈ [38,63].
- We used **Sentinel-2 images**, excluding S1 and SPOT-6 sensors.
- We explored two expressions of data scarcity:
  - **Label availability**: 1, 10, 50, 100%.
  - **Temporal resolution**: 1, 6, 15, 25, 35 instances

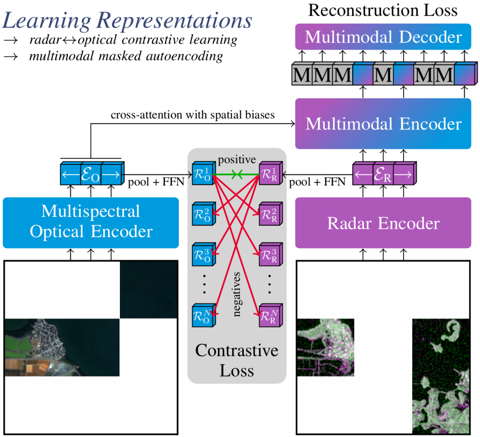


Methodology: Encoder and Decoder Models

- Two GFMs were tested as **encoders**:

Methodology: Encoder and Decoder Models

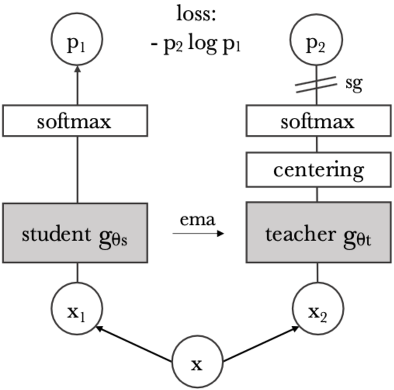
- Two GFMs were tested as **encoders**:
  - CROMA.



Fuller, A. et al. "CROMA: Remote Sensing Representations with Contrastive Radar-Optical Masked Autoencoders." Advances in Neural Information Processing Systems, 36(NeurIPS):1–33, 2023.

Methodology: Encoder and Decoder Models

- Two GFMs were tested as **encoders**:
  - CROMA.
  - SSL4EO-DINO.



Caron, M. et al. "Emerging Properties in Self-Supervised Vision Transformers." IEEE/CVF International Conference on Computer Vision (ICCV), 2018.

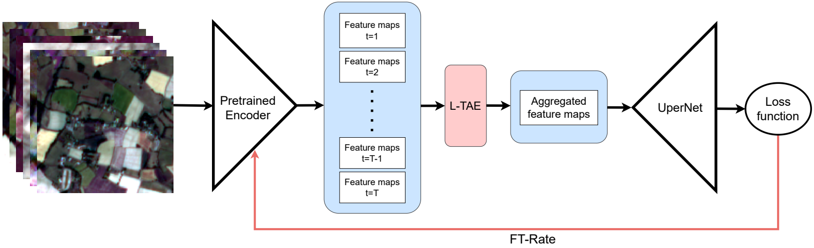
Wang, Y., et al. "SSL4EO-S12: A Large-Scale Multi-Modal, Multi-Temporal Dataset for Self-Supervised Learning in Earth Observation." IEEE Geoscience and Remote Sensing Magazine, Vol. 11, Issue 3, pp. 98 - 106, 2023.

Methodology: Encoder and Decoder Models

- Two GFMs were tested as **encoders**:
  - CROMA.
  - SSL4EO-DINO.
- An UperNet* was adopted for the **decoder**.

Fine-Tuning Policy

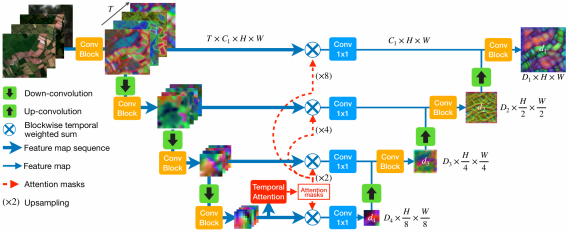
$$\text{FT-Rate} = \frac{lr_{\text{encoder}}}{lr_{\text{decoder}}}$$



*Xiao, T., et al. "Unified Perceptual Parsing for Scene Understanding." European Conference on Computer Vision (ECCV), 2018.

Methodology: Encoder and Decoder Models

- Two GFMs were tested as **encoders**:
  - CROMA.
  - SSL4EO-DINO.
- An UperNet* was adopted for the **decoder**.
- We compare GFMs to a **baseline**: fully-supervised U-TAE.



Garnot, V.S.F., et al. "Panoptic Segmentation of Satellite Image Time Series with Convolutional Temporal Attention Networks." Proceedings of the IEEE International Conference on Computer Vision, pp. 4852–4861, 2021.

Results: Abundant labels

- U-TAE is always better than frozen encoders when t $>$ 1.

Model	FT-Rate	100% of data					50% of data				
		1	6	15	25	35	1	6	15	25	35
CROMA	0.0	16.10	43.40	53.37	55.66	56.71	14.83	39.83	47.43	51.02	52.15
	0.1	19.52	49.65	58.11	60.41	61.12	17.44	45.54	53.08	56.95	57.50
	1.0	17.71	49.15	56.55	60.12	61.06	15.71	45.03	52.71	56.20	57.10
DINO	0.0	15.55	38.08	47.41	49.59	50.68	13.52	34.55	42.52	44.30	44.92
	0.1	16.43	42.50	51.17	54.03	55.19	14.36	37.34	46.98	48.65	50.24
	1.0	16.91	44.61	53.65	56.14	57.23	15.34	40.28	49.16	51.85	53.02
U-TAE	-	13.75	**45.45**	**54.33**	**57.36**	**58.98**	11.52	**40.99**	**51.50**	**53.85**	**55.18**

Results: Abundant labels

- U-TAE is always better than frozen encoders when t $>$ 1.
- CROMA with partial fine-tuning reported the best results.

Model	FT-Rate	100% of data					50% of data				
		1	6	15	25	35	1	6	15	25	35
CROMA	0.0	16.10	43.40	53.37	55.66	56.71	14.83	39.83	47.43	51.02	52.15
	0.1	**19.52**	**49.65**	**58.11**	**60.41**	**61.12**	**17.44**	**45.54**	**53.08**	**56.95**	**57.50**
	1.0	17.71	49.15	56.55	60.12	61.06	15.71	45.03	52.71	56.20	57.10
DINO	0.0	15.55	38.08	47.41	49.59	50.68	13.52	34.55	42.52	44.30	44.92
	0.1	16.43	42.50	51.17	54.03	55.19	14.36	37.34	46.98	48.65	50.24
	1.0	16.91	44.61	53.65	56.14	57.23	15.34	40.28	49.16	51.85	53.02
U-TAE	-	13.75	45.45	54.33	57.36	58.98	11.52	40.99	51.50	53.85	55.18

Results: Scarce labels

- CROMA with frozen encoder outperforms U-TAE.

Model	FT-Rate	10% of data					1% of data				
		1	6	15	25	35	1	6	15	25	35
CROMA	**0.0**	**10.89**	26.17	34.05	36.28	37.37	**7.21**	16.62	**20.64**	**22.77**	**25.53**
	0.1	13.09	30.07	37.89	39.43	40.67	7.30	18.63	22.81	23.26	24.06
	1.0	12.47	32.73	40.58	42.29	45.30	6.44	20.81	24.46	25.64	26.42
DINO	0.0	10.23	23.18	28.64	31.11	31.72	6.13	13.69	16.53	17.30	15.99
	0.1	11.00	25.06	32.83	35.79	36.09	6.52	14.56	17.69	19.12	18.71
	1.0	11.03	27.75	35.60	37.94	39.56	6.98	17.38	20.29	21.77	22.38
U-TAE	-	9.24	28.66	37.33	38.68	39.44	5.43	18.96	19.45	19.62	21.40

Results: Scarce labels

- CROMA with frozen encoder outperforms U-TAE.
- CROMA with full fine-tuning reported the best results.

Model	FT-Rate	10% of data					1% of data				
		1	6	15	25	35	1	6	15	25	35
CROMA	0.0	10.89	26.17	34.05	36.28	37.37	7.21	16.62	20.64	22.77	25.53
	0.1	**13.09**	30.07	37.89	39.43	40.67	**7.30**	18.63	22.81	23.26	24.06
	1.0	12.47	**32.73**	**40.58**	**42.29**	**45.30**	6.44	**20.81**	**24.46**	**25.64**	**26.42**
DINO	0.0	10.23	23.18	28.64	31.11	31.72	6.13	13.69	16.53	17.30	15.99
	0.1	11.00	25.06	32.83	35.79	36.09	6.52	14.56	17.69	19.12	18.71
	1.0	11.03	27.75	35.60	37.94	39.56	6.98	17.38	20.29	21.77	22.38
U-TAE	-	9.24	28.66	37.33	38.68	39.44	5.43	18.96	19.45	19.62	21.40

Results: Scarce labels

- CROMA with frozen encoder outperforms U-TAE.
- CROMA with full fine-tuning reported the best results.
- Full fine-tuned DINO outperforms U-TAE.

Model	FT-Rate	10% of data					1% of data				
		1	6	15	25	35	1	6	15	25	35
CROMA	0.0	10.89	26.17	34.05	36.28	37.37	7.21	16.62	20.64	22.77	25.53
	0.1	13.09	30.07	37.89	39.43	40.67	7.30	18.63	22.81	23.26	24.06
	1.0	12.47	32.73	40.58	42.29	45.30	6.44	20.81	24.46	25.64	26.42
DINO	0.0	10.23	23.18	28.64	31.11	31.72	6.13	13.69	16.53	17.30	15.99
	0.1	11.00	25.06	32.83	35.79	36.09	6.52	14.56	17.69	19.12	18.71
	1.0	**11.03**	27.75	35.60	37.94	**39.56**	**6.98**	17.38	**20.29**	**21.77**	**22.38**
U-TAE	-	9.24	28.66	37.33	38.68	39.44	5.43	18.96	19.45	19.62	21.40

Conclusions

- In high-label regimes: task-specific models can compete with, but are not better than, fine-tuned GFMs **(and are cheaper)**.

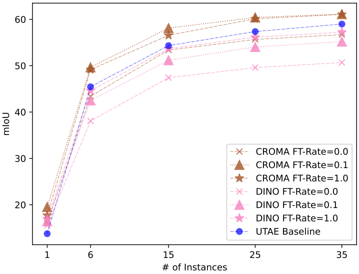
Model	# of Trainable Parameters (M)	
	Only Decoder	Whole network
CROMA	46.95	350.0
DINO	30.89	53.5
U-TAE	-	1.1

Conclusions

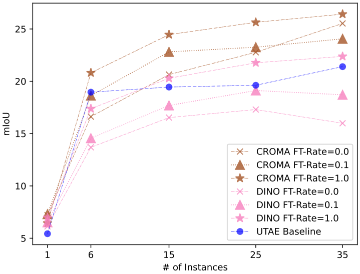
- In high-label regimes: task-specific models can compete with, but are not better than, fine-tuned GFMs **(and are cheaper)**.
- Careful fine-tuning is key to leverage GFMs:

Conclusions

- In high-label regimes: task-specific models can compete with, but are not better than, fine-tuned GFMs **(and are cheaper)**.
- Careful fine-tuning is key to leverage GFMs:
  - Treat FT-Rate as a relevant hyperparameter.



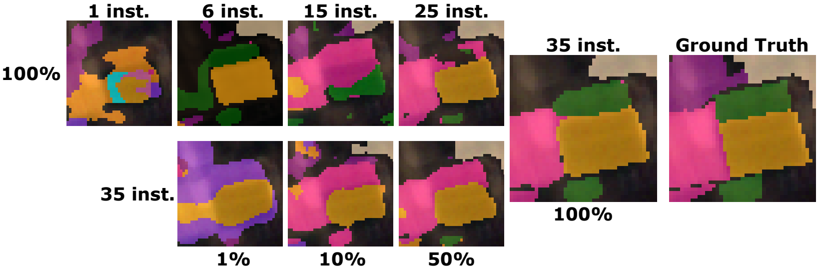
(a) 100% of data.



(b) 1% of data.

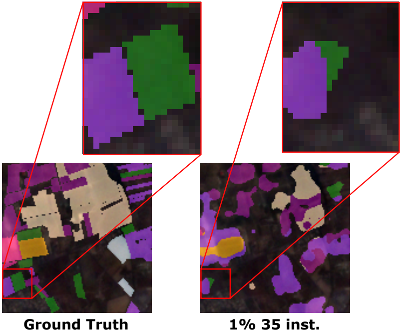
Conclusions

- In high-label regimes: task-specific models can compete with, but are not better than, fine-tuned GFMs **(and are cheaper)**.
- Careful fine-tuning is key to leverage GFMs:
  - Treat FT-Rate as a relevant hyperparameter.
  - Temporal resolution and label availability affect performance.



Conclusions

- In high-label regimes: task-specific models can compete with, but are not better than, fine-tuned GFMs **(and are cheaper)**.
- Careful fine-tuning is key to leverage GFMs:
  - Treat FT-Rate as a relevant hyperparameter.
  - Temporal resolution and label availability affect performance.
- Performance improves with longer time-sequences in low-label settings, even for frozen GFMs.



Thank you for your attention!

Contact:

giovanni.castiglioni@cenia.cl