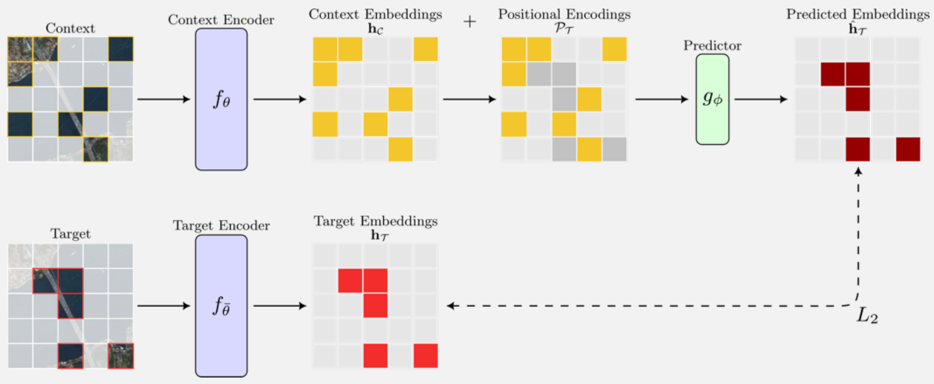
ECEO-EPFL

Joint-Embedding Predictive Architectures with Domain-Aware Masking for Foundation Model Pretraining in Earth Observation

Ozan Arda Güven, Gencer Sumbul, Devis Tuia
Environmental Computational Science and Earth Observation Laboratory (ECEO), EPFL
Contact: gencer.sumbul@epfl.ch

Motivation and Aim

- Foundation models based on Joint-Embedding Predictive Architectures (I-JEPA) [1] are effective at capturing rich semantic structures with minimal labelling, but their adaptation to Earth Observation (EO) remains challenging due to the complex nature of satellite imagery.



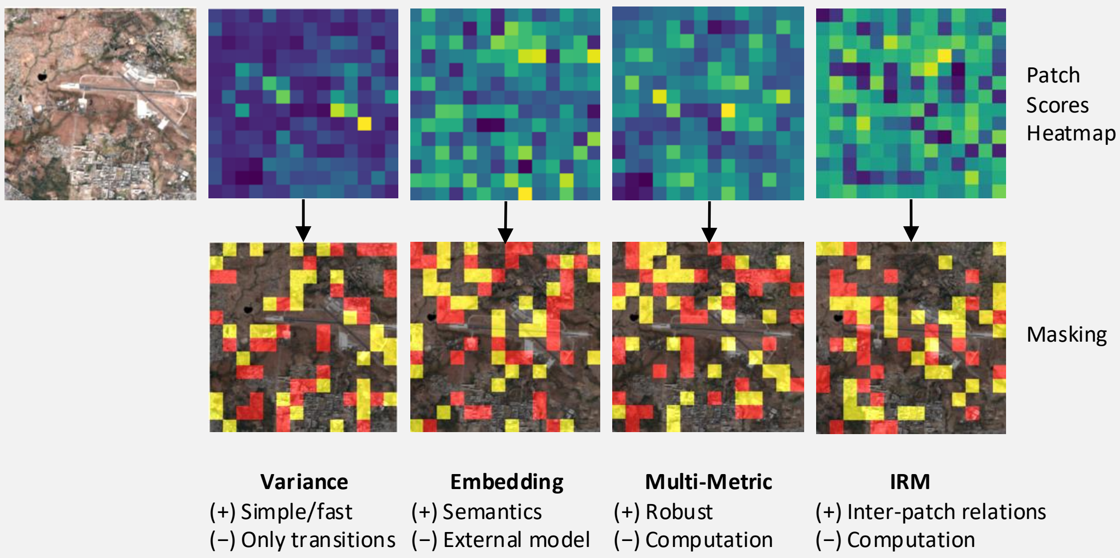
- I-JEPA, with its multi-block strategy, maintains spatial coherence for masking with random rectangular blocks. For satellite imagery:
  - Key features (urban structures, roads) might be small or sparse
  - Large homogeneous areas (desert, ocean) are less informative
  - Contiguous big blocks miss subtle transitions (critical for EO)

- For EO, domain-aware masking strategies are needed based on semantic importance rather than purely random blocks.

Methodology

- We propose four domain-aware patch-centric masking strategies, selecting patches to mask with high semantic significance based on different patch-level scoring schemes.

1) Variance Masking scores patches based on the variance of pixel intensities in each patch.
2) Embedding Masking scores patches based on distance across patch embeddings obtained from a pre-trained model.
3) Multi-Metric Masking scores patches based on the combination of variance, embedding distance, entropy, and edge density.
4) Informativeness–Representativeness Masking (IRM) considers inter-patch relations based on relevancy, hardness and diversity.



Results

- We pretrained ViT-B and ViT-L based models with 100K RGB images from Functional Map of the World (fMoW) [2]. We applied k-NN (k=20), linear probing (LP), and finetuning (FT) for scene classification on EuroSAT [3] and RESISC-45 [4].

Masking Strategy		ViT-B		ViT-L	
		k-NN	LP	k-NN	LP
I-JEPA [1]	Multi-Block	95.7	97.7	93.1	97.0
Ours	Variance	95.6	97.5	95.0	97.4
	Embedding	96.3	**97.8**	95.1	97.7
	Multi-Metric	**96.4**	**97.8**	94.9	97.4
	IRM	96.2	97.2	**95.8**	**97.9**

Scene Classification Results on EuroSAT

Masking Strategy		ViT-B		ViT-L	
		LP	FT	LP	FT
I-JEPA [1]	Multi-Block	88.4	85.8	88.4	82.7
Ours	Multi-Metric	**89.9**	87.0	89.0	84.4
	IRM	**89.9**	**87.2**	**89.5**	**86.0**

Scene Classification Results on RESISC-45

Conclusion and Discussion

- Proposed domain-aware masking strategies not only outperform random multi-block masking for I-JEPA, but also allow encoding the EO-specific features more effectively.
- Model capacity and pretraining significantly influence the efficiency of the proposed strategies: smaller backbones converge faster, yet larger ones benefit from longer schedules.
- Our findings underscore the potential for adapting foundation models to complex EO data, especially by using domain-aware masking strategies for not only I-JEPA but also MAE.

References

[1] M. Assran, Q. Duval, I. Misra, P. Bojanowski, P. Vincent, M. Rabbat, Y. LeCun, and N. Ballas, “Self-supervised learning from images with a joint-embedding predictive architecture,” IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 2334–2344.
[2] G. Christie, N. Fendley, J. Wilson, and R. Mukherjee, “Functional Map of the World,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6172–6180.
[3] P. Helber, B. Bischke, A. Dengel, and D. Borth, "EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification,“ IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019.
[4] G. Cheng, J. Han, and X. Lu, “Remote sensing image scene classification: Benchmark and state of the art,” Proceedings of the IEEE, 105(10), 1865–1883, 2017.

This work is supported by the European Space Agency through the Discovery and Preparation Program, and is part of the project “Toward a Foundation Model for Multi-Sensor Earth Observation Data with Language Semantics” (ESA Contract no. 4000144146).