Harnessing Multi-Modal Co-learning for Missing Earth Observation Modalities

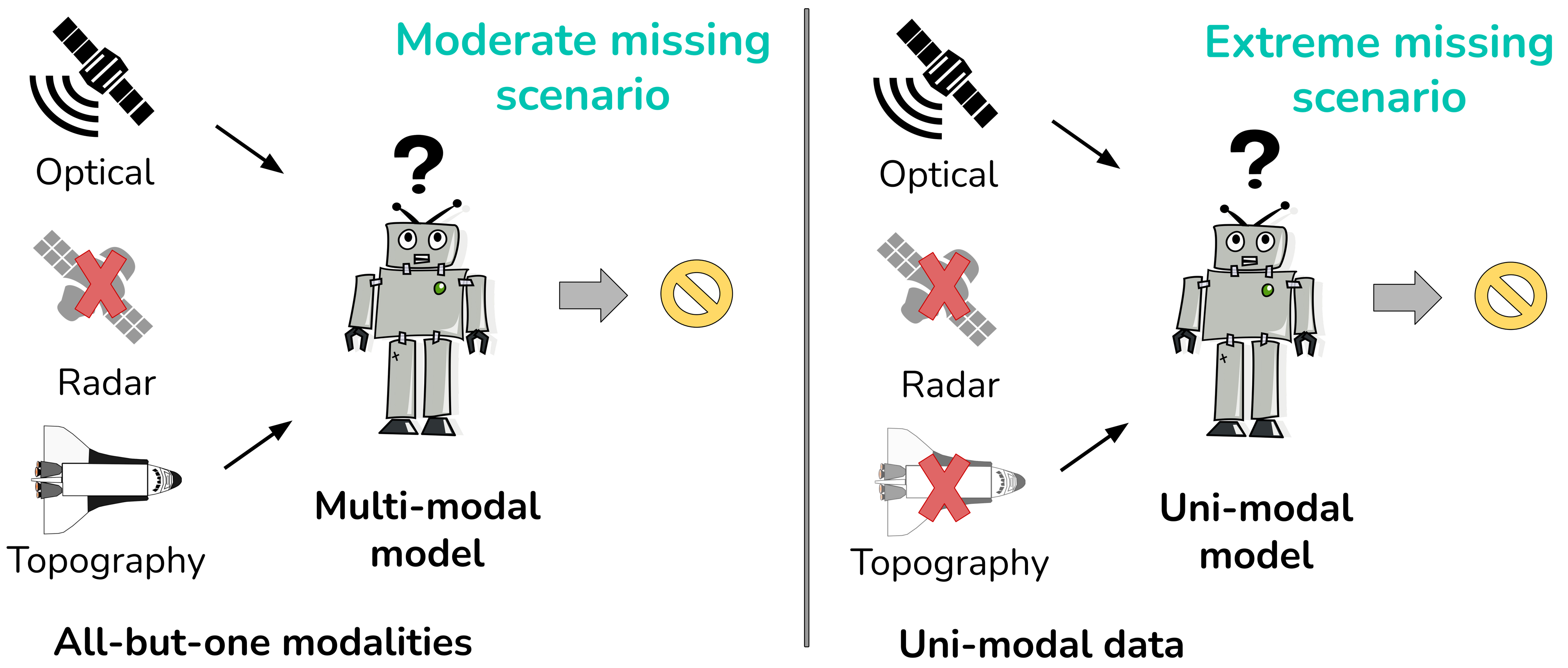
Francisco Mena, Dino Ienco, Cássio F. Dantas, Roberto Interdonato

GFZ Helmholtz Centre for Geoscience, University of Kaiserslautern-Landau
INRAE/CIRAD UMR TETIS, University of Montpellier, EVERGREEN (Inria)

GFZ RPTU FONDA Inria

Motivation

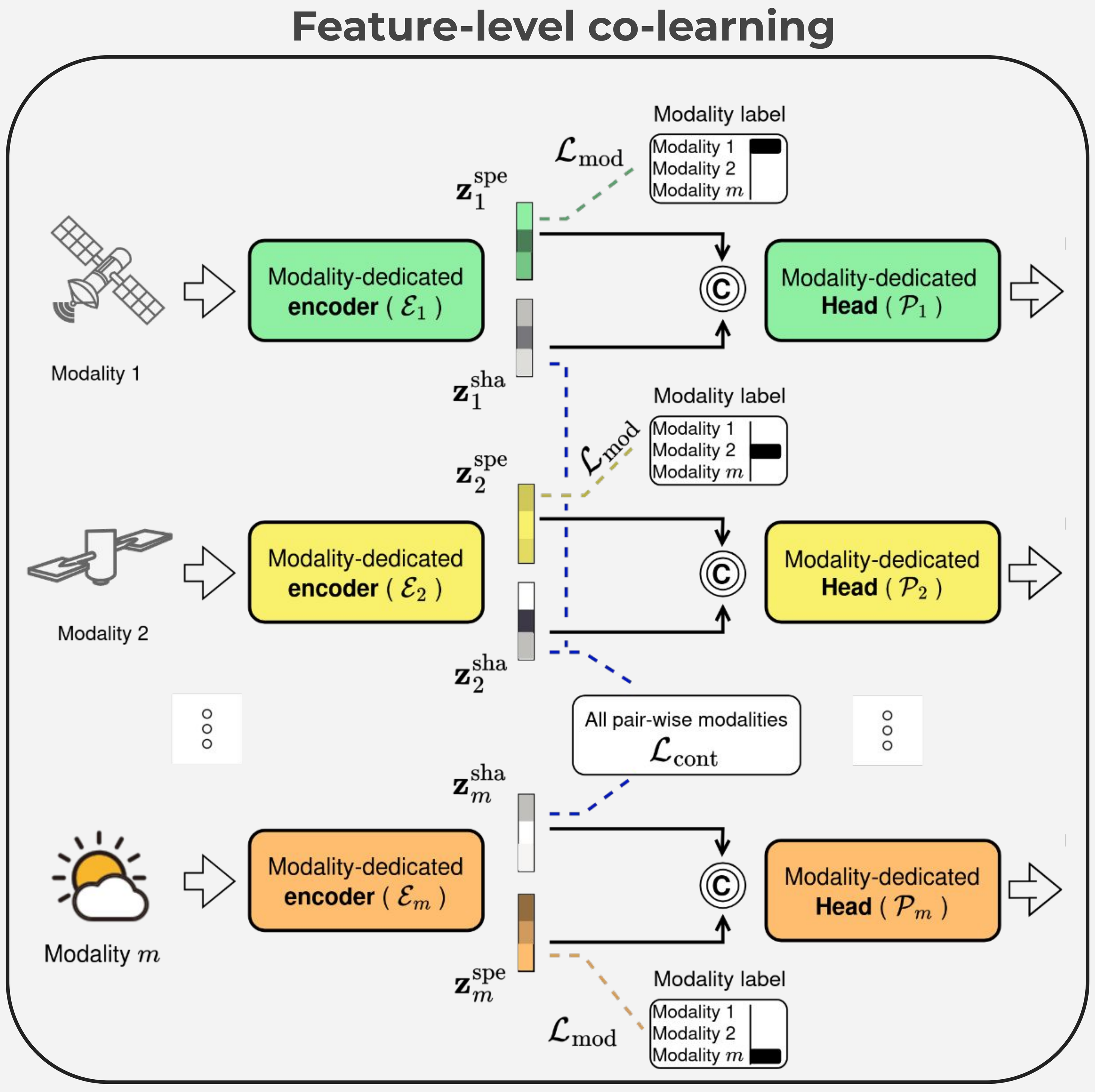
- Context: **Multi-modal sensor data** has allowed effective model mapping for diverse Earth Observation (EO) applications, such as crop-type classification.
- Problem: Various factors can affect the **consistent spatio-temporal availability** of EO modalities [1]: sensor malfunctions, spectral noise, cloud cover, or limited coverage.
- Goal: Maintain the predictive performance of multi-modal models **(robustness)** when **arbitrary** modalities are available at **inference time** (e.g. in moderate and extreme missingness).



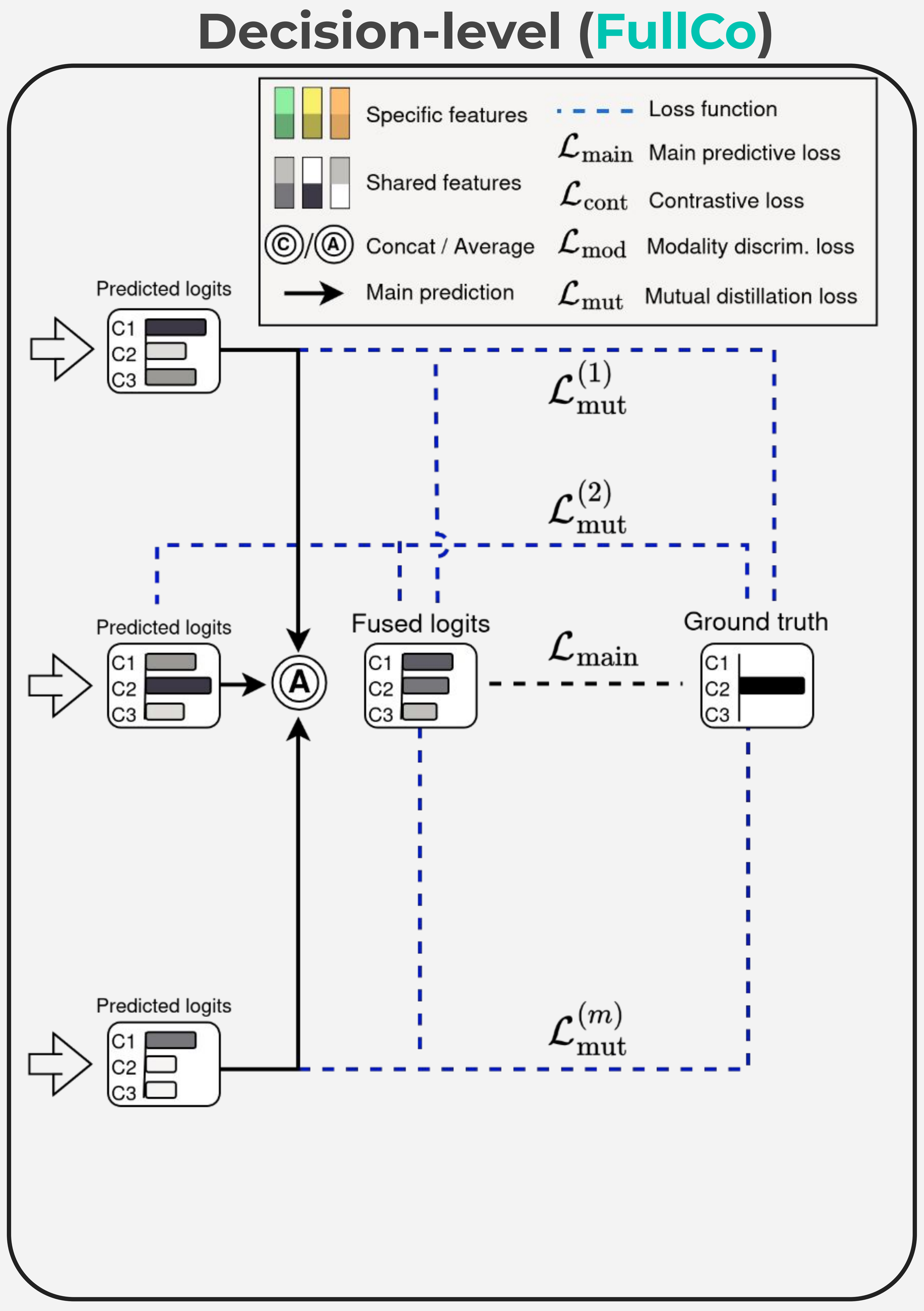
Methodology

- **Main premise**: Modality-dedicated models should collaborate with each other (**co-learning** [2]) at the feature and decision levels.

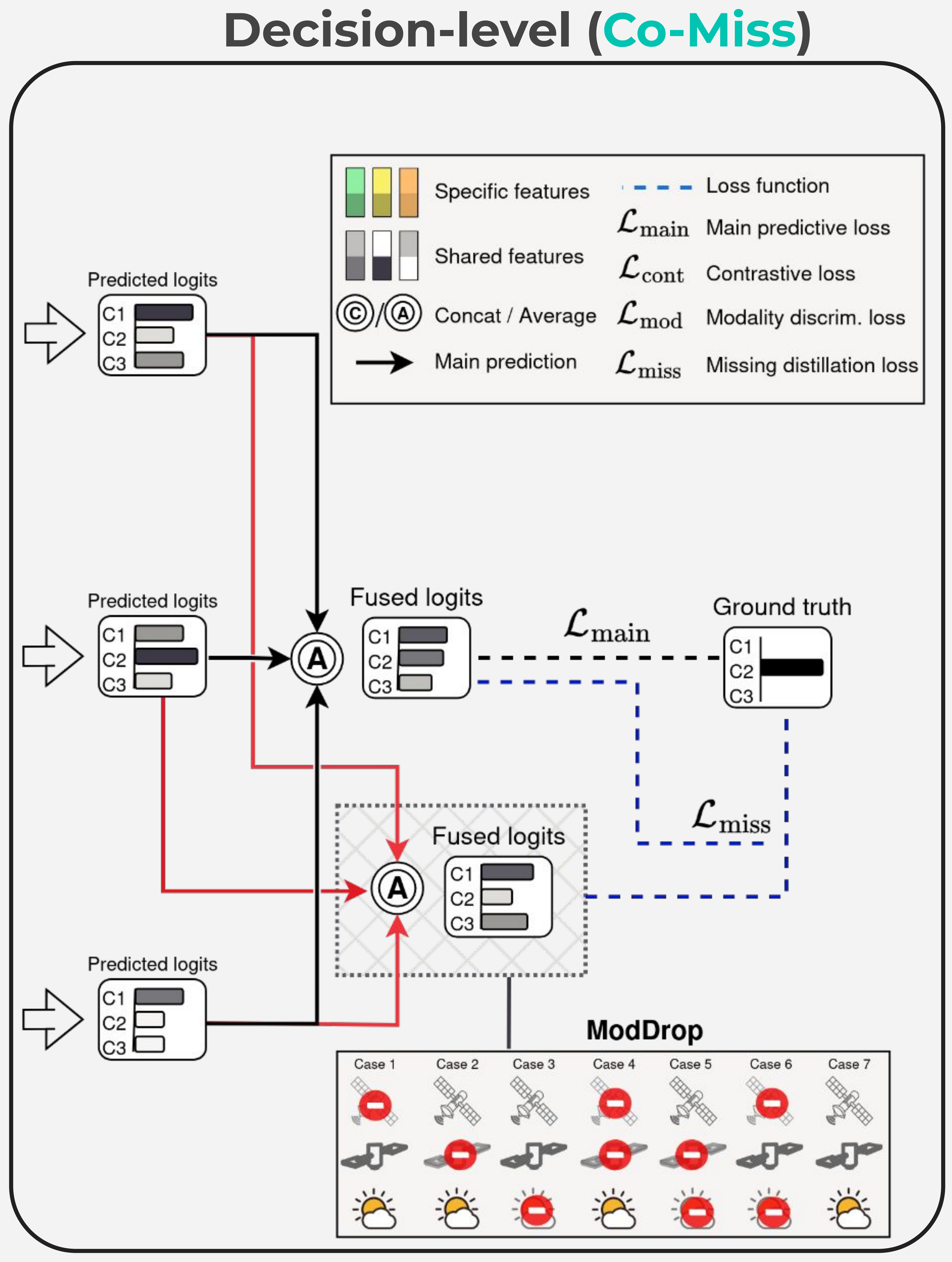
Feature-level co-learning



Decision-level (FullCo)



Decision-level (Co-Miss)



- Multiple loss functions to optimize:
  - *Modality discriminant:* to guide the **specific** features.
  - *Cross-modal contrastive*: to guide the **shared** features.
  - *Main predictive*: to guide the **fused** (full-modal) prediction.
- FullCo (Method variant 1)
  - **Mutual distillation**: to guide the **individual** (modality-dedicated) predictions.
- Co-Miss (Method variant 2)
  - **Missing distillation**: to guide the prediction fused with random **missing modalities.**

Validation Dataset

- CropHarvest: Pixel-wise **crop classification** (10 classes).
- Four modalities:
  - **Optical:** from Sentinel-2 multi-spectral data.
  - **Radar:** from Sentinel-1 data.
  - **Weather:** from ERA5 data.
  - **Topography:** from NASA's SRTM data.

Opt.	Rad.	Wea.	Top.	Uni - modal	AnySat [3]	Galileo [4]	FCoM-av [5]	TiMML [6]	DSensD+ [7]	FullCo	Co-Miss
●	●	●	●		71.2	69.7	76.7	72.4	76.5	76.1	**78.0**
○	●	●	●				66.6	63.1	65.9	66.1	**66.9**
●	○	●	●				74.6	71.0	75.4	75.1	**76.9**
●	●	○	●				76.2	72.7	76.4	76.3	**77.7**
●	●	●	○				76.5	72.5	76.8	76.3	**78.0**
Moderate average							73.5	70.0	73.6	73.4	**74.9**
●	○	○	○	71.0	70.9	70.5	73.8	71.5	75.3	76.0	**76.3**
○	●	○	○	56.0	51.5	50.4	53.6	53.6	57.3	**59.2**	57.9
○	○	●	○	46.7		48.3	42.8	44.3	48.4	**50.2**	48.8
○	○	○	●	28.0		30.1	19.1	2.6	31.5	**31.7**	30.7
Extreme average				50.4	61.2*	49.8*	47.3	43.0	53.1	**54.3**	53.4
Overall average				50.4*	64.6*	53.8*	62.2	58.2	64.9	65.0	**65.4**

F1 scores, * Averaged over available cases.

Findings

★ FullCo method is better for **extreme** missing conditions.

★ Co-Miss method is more effective in **moderate** missingness, and in average.

Future work:

- Validate in other datasets.
- Ablation (impact of individual losses and components).

References

[1] Shen, Huanfeng, et al. "Missing information reconstruction of remote sensing data: A technical review." IEEE Geoscience and Remote Sensing Magazine 3.3 (2015): 61-85.
[2] Kieu, Nhi, et al. "Multimodal colearning meets remote sensing: Taxonomy, state of the art, and future works." IEEE JSTARS 17 (2024): 7386-7409.
[3] Astruc, Guillaume, et al. "AnySat: One Earth Observation Model for Many Resolutions, Scales, and Modalities." Proceedings of the CVPR conference. 2025.
[4] Tseng, Gabriel, et al. "Galileo: Learning Global & Local Features of Many Remote Sensing Modalities." arXiv preprint arXiv:2502.09356 (2025).
[5] Mena, Francisco, et al. "Missing data as augmentation in the Earth observation domain: A multi-view learning approach." Neurocomputing 638 (2025): 130175.
[6] Xu, Guozheng, et al. "Transformer-based incomplete multi-modal learning for land cover classification." IGARSS 2024-2024 IEEE IGARSS. IEEE, 2024.
[7] Mena, Francisco, et al. "Multi-sensor Model for Earth Observation Robust to Missing Data via Sensor Dropout and Mutual Distillation." IEEE Access (2025).