Generative Domain Adaptation and Foundation Models Benchmarking for Robust Earth Observation

Georges Le Bellier
@lebellig.bsky.social
Conservatoire National des Arts et Métiers

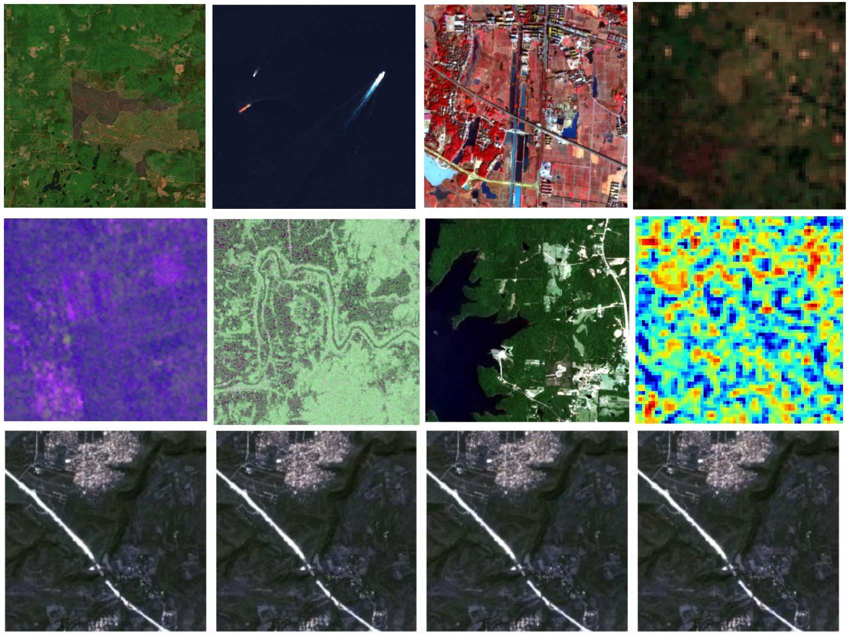
Learning Machines Seminar

Introduction

EO: rich and diverse source of data

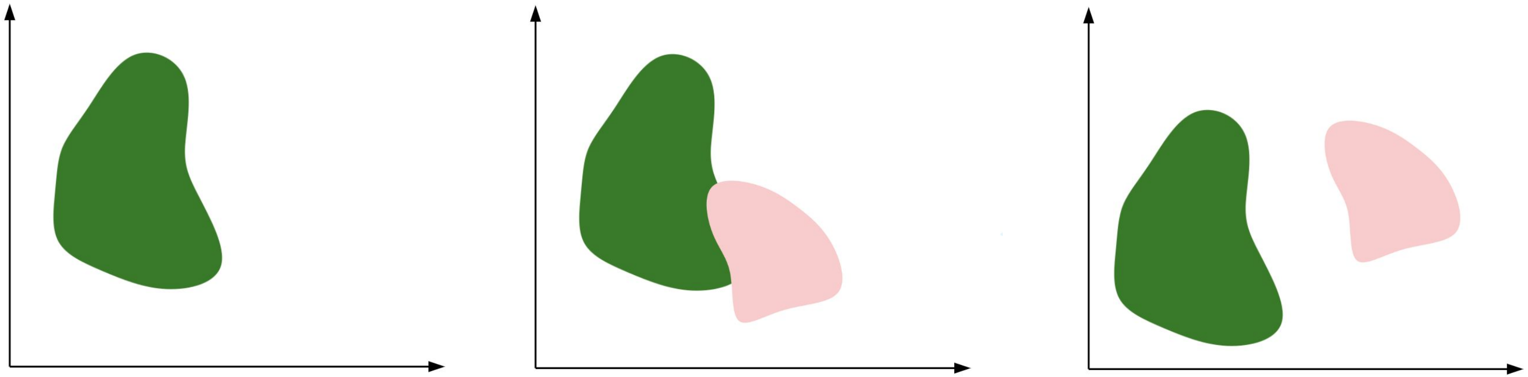
Offers differents views on Earth systems, environmental change, phenomenon occurring on Earth.

Challenging topic for deep learning



Introduction

distribution mismatch ⇒ poor performances



Training distribution

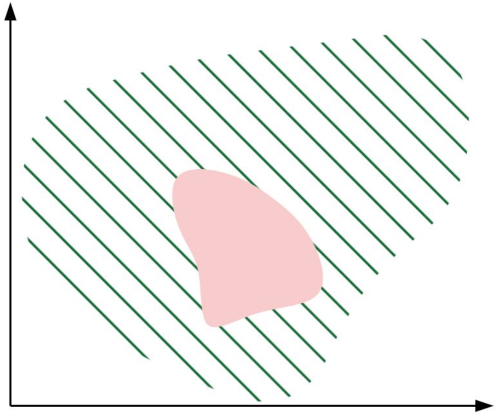
Training distribution
Inference distribution
(e.g. diverse locations...)

Training distribution
Inference distribution
(e.g. different sensors...)

Introduction

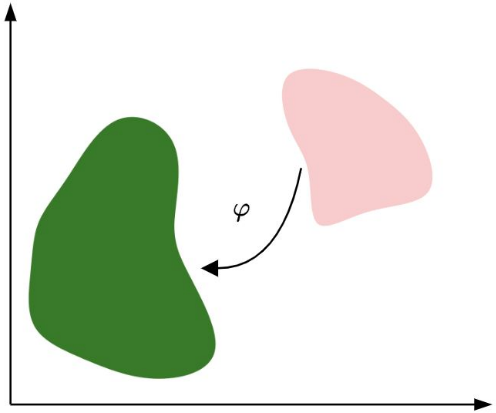
Our goal is for EO pipelines to be robust against distribution shifts (sensors, areas, natural disasters...)

1. Geospatial Foundation Models (GFMs)



Large pretraining distribution 
→ learn rich image representations

2. Generative Domain Adaptation



→ Transport the inference distribution to the training one

PLAN

Part.1. PANGAEA

- Geospatial Foundation Models
- PANGAEA framework
- Evaluation Protocol
- Results
- Conclusion

Part.2. FlowEO

- Domain adaptation
- Flow Matching
- FlowEO
- Experiments & Results
- Conclusion

Part.1

PANGAEA: a Global and Inclusive Benchmark for Geospatial Foundation Models

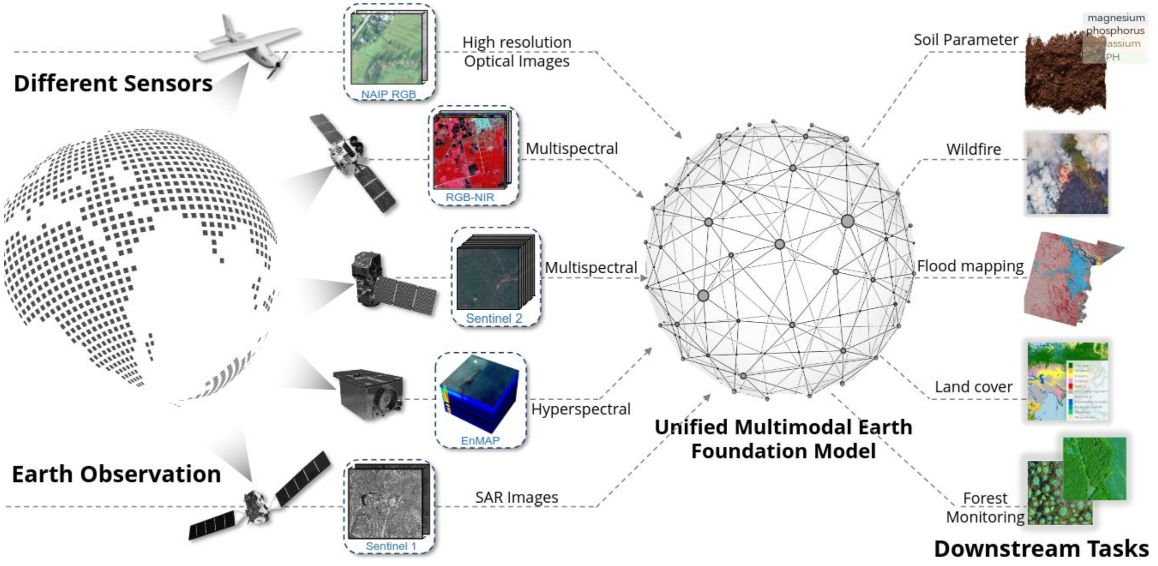
Valerio Marsocci*, Yuru Jia*, Georges Le Bellier, David Kerekes, Liang Zeng, Sebastian Hafner, Sebastian Gerard, Eric Brune, Ritu Yadav, Ali Shibli, Heng Fang, Yifang Ban, Maarten Vergauwen, Nicolas Audebert, Andrea Nascetti

PLAN

Geospatial Foundation Models

Goal: learn useful representations of EO images on large datasets

- task agnosticism
- spatio-temporal awareness
- sensor agnosticism
- multimodality
- adaptability



source: DOFA

Geospatial Foundation Models

Large list of GFMs: which one should I use to solve my problem ?

We need a robust benchmark

- GFMs evaluate themselves on different setups
- Benchmark targets:

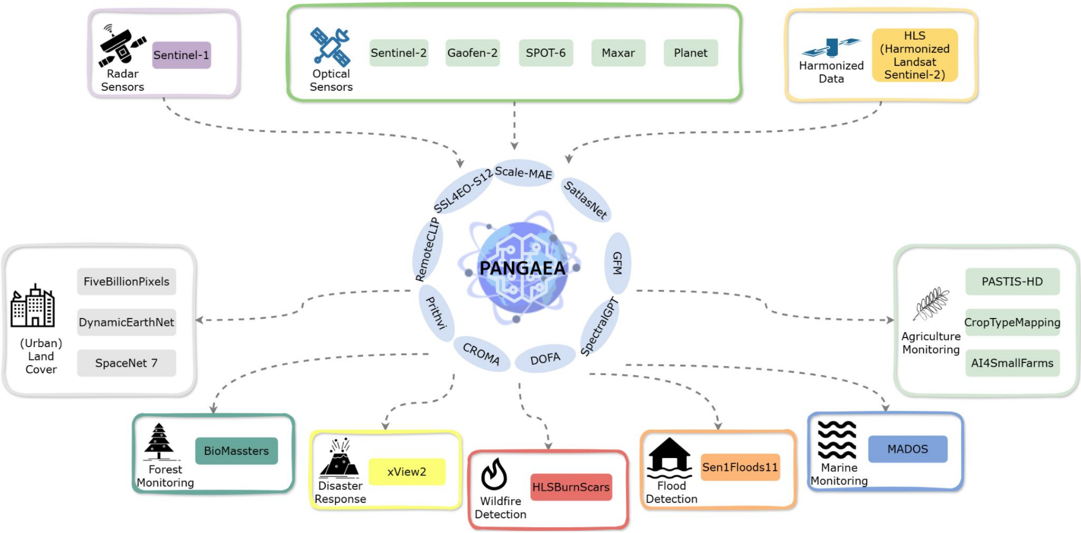
1. Performance evaluation
2. Fairness and robustness
3. Guide improvements

PLAN

Part.1. PANGAEA
- Geospatial Foundation Models
- PANGAEA framework
- Evaluation Protocol
- Results
- Conclusion

Part.2. FlowEO
- Domain adaptation
- Flow Matching
- FlowEO
- Experiments & Results
- Conclusion

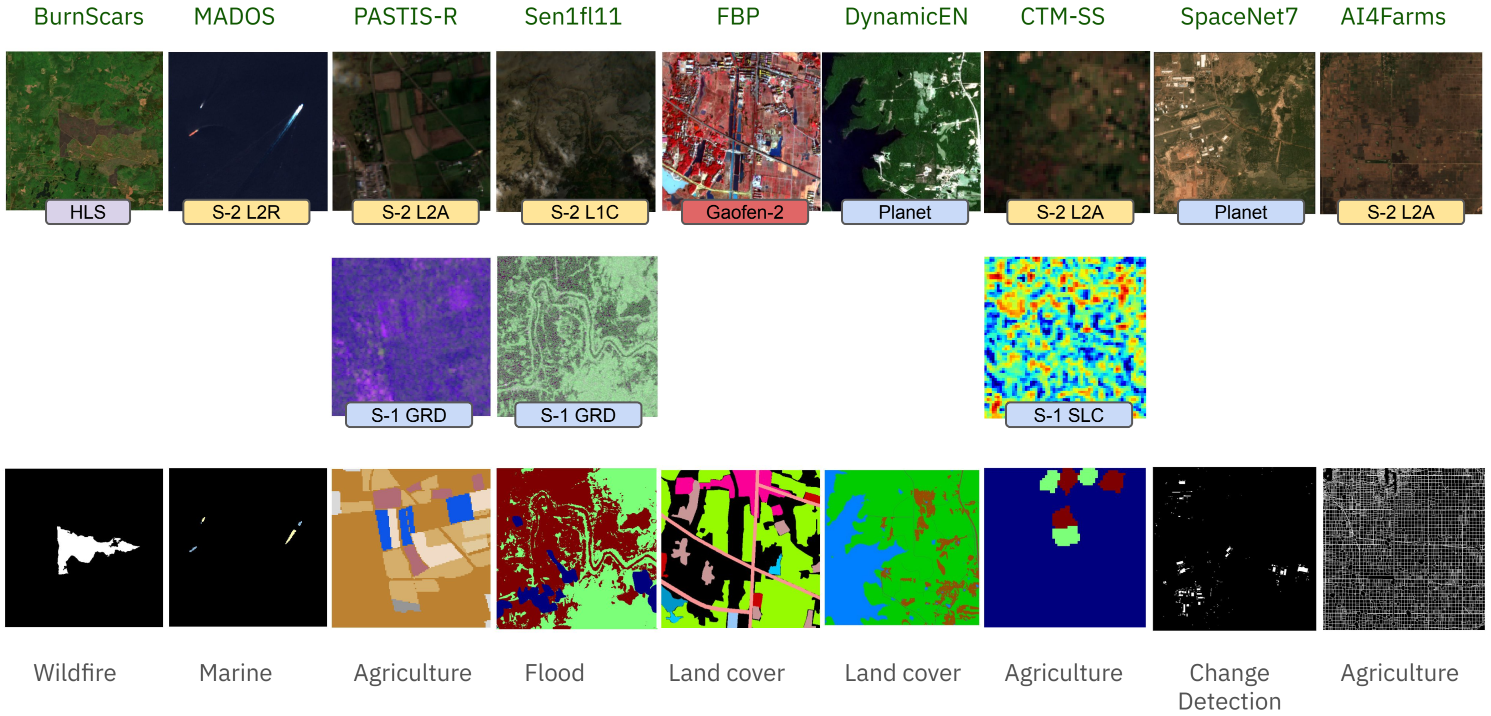
PANGAEA framework



PANGAEA research questions:

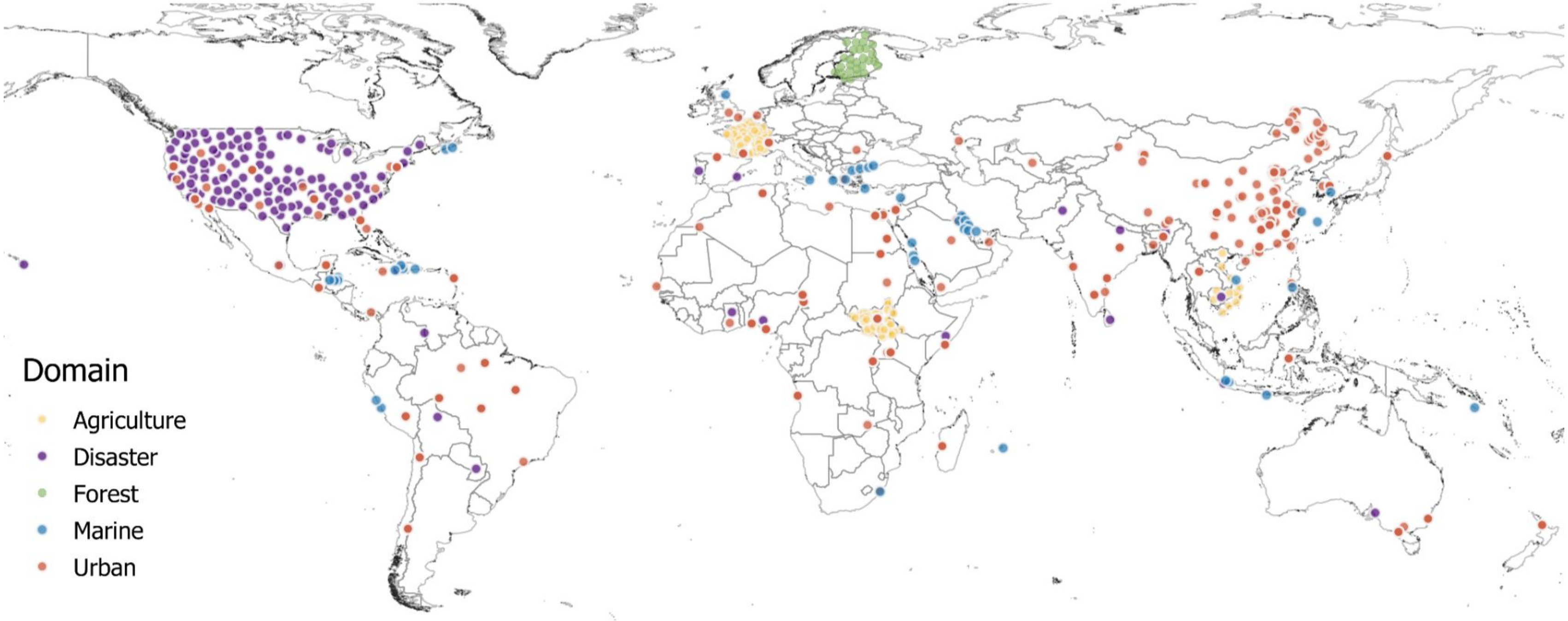
1. Generalization across domains
2. Comparison to supervised baselines
3. Performance with limited labels

Datasets



→ Pixel-level prediction (segmentation/regression/change detection)

Datasets



Geospatial Foundation Models

Selection criterions:
– Open source GFMs + publicly available weigths (reproducibility)
– Impact on the community

Model	Pretraining Images	Patches/Volume
CROMA	Sentinel-1, Sentinel-2	3M
DOFA	Sentinel-1, Sentinel2, Gaofen-2, NAIP, EnMAP	8.08M
GFM-Swin	NAIP, RSD46-WHU, MLRSNet, RESISC45, PatternNet	600K
Prithvi	Harmonized Landsat Sentinel-2 (HLS)	1TB
RemoteCLIP	SEG-4, DET-10, RET-3	165K
SatlasNet	Sentinel-2, NAIP	856K
Scale-MAE	FMoW-RGB	363.6K
SpectralGPT	fMoW-S2, BigEarthNet	1.47M
SSL4EO-S12	Sentinel-1, Sentinel-2	3M

PLAN

Part.1. PANGAEA
- Geospatial Foundation Models
- PANGAEA framework
- Evaluation Protocol
- Results
- Conclusion

Part.2. FlowEO
- Domain adaptation
- Flow Matching
- FlowEO
- Experiments & Results
- Conclusion

Evaluation protocol

1. Dataset preprocessing

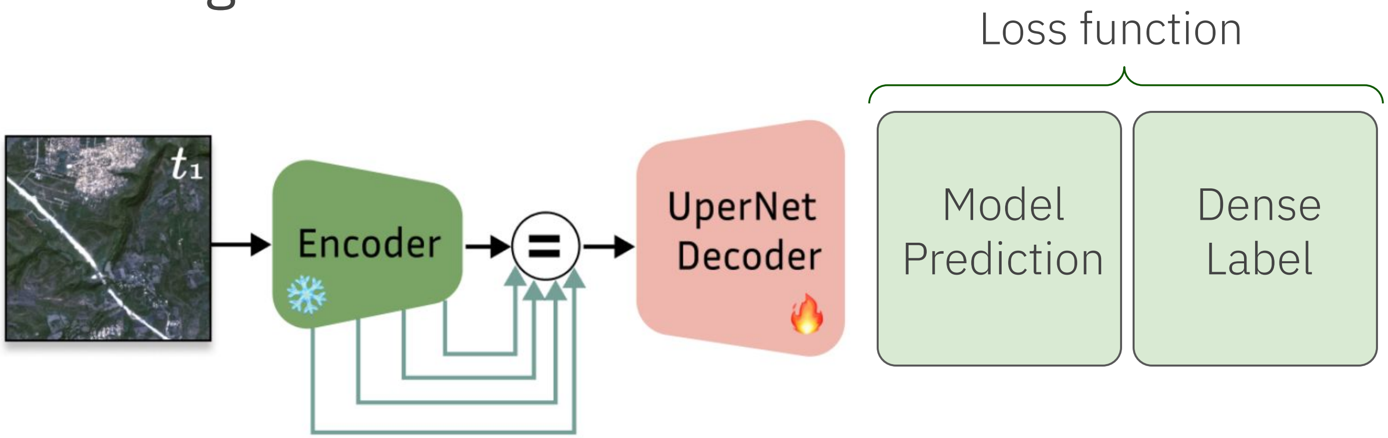
- Same dataset preprocessing for all models (band-wise normalization)

We need to match datasets available bands with GFMs input bands

- Band Matching + Adaptation (corresponding bands)
- Zero-padding for missing bands

Evaluation protocol

2. Decoder training

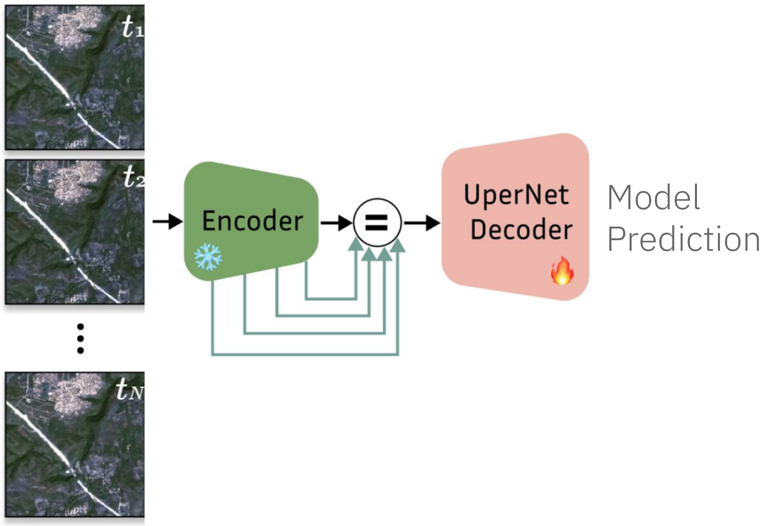


- GFM are frozen encoders (usable for everyone)
- UperNet decoder

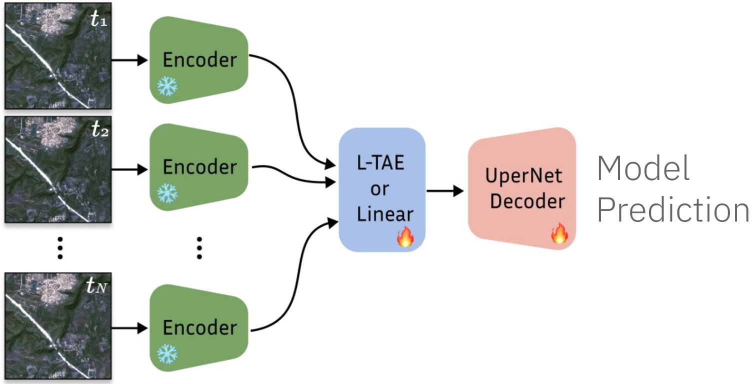
Evaluation protocol

3. Multi-temporal datasets

Multi-temporal models (e.g. Prithvi)



Single-temporal models



Two different temporal aggregation strategies: linear or L-TAE

Evaluation protocol

4. Data-scarcity
   - 50% or 10% of labels to train the decoder

5. Supervised baselines
   - Two supervised baselines: UNet and ViT-B/16

PLAN

Part.1. PANGAEA

- Geospatial Foundation Models
- PANGAEA framework
- Evaluation Protocol
- Results
- Conclusion

Part.2. FlowEO

- Domain adaptation
- Flow Matching
- FlowEO
- Experiments & Results
- Conclusion

Main results

Table 5: Performance evaluation of Geospatial Foundation Models across 11 benchmark datasets using 100% of the data. For semantic segmentation and change detection tasks, the mIoU ↑ is reported. For regression task, RMSE ↓ is reported. #Top2 indicates the number of datasets where the models achieve top-2 performance across all evaluated datasets.

Model	HLS Burns	MADOS	PASTIS	Sen1Floods11	xView2	FBP	DynEarthNet	CropMap	SN7	AI4Farms	BioMassters	#Top2
CROMA	82.42	**67.55**	32.32	90.89	53.27	51.83	38.29	49.38	59.28	25.65	36.81	2
DOFA	80.63	59.58	30.02	89.37	59.64	43.18	39.29	51.33	61.84	27.07	42.81	2
GFM-Swin	76.90	64.71	21.24	72.60	59.15	67.18	34.09	46.98	60.89	27.19	46.83	1
Prithvi	83.62	49.98	33.93	90.37	49.35	46.81	27.86	43.07	56.54	26.86	39.99	1
RemoteCLIP	76.59	60.00	18.23	74.26	57.41	**69.19**	31.78	52.05	57.76	25.12	49.79	2
SatlasNet	79.96	55.86	17.51	90.30	52.23	50.97	36.31	46.97	61.88	25.13	41.67	0
Scale-MAE	76.68	57.32	24.55	74.13	**60.72**	67.19	35.11	25.42	**62.96**	21.47	47.15	3
SpectralGPT	80.47	57.99	35.44	89.07	48.40	33.42	37.85	46.95	58.86	26.75	36.11	1
S12-MoCo	81.58	51.76	34.49	89.26	51.59	53.02	35.44	48.58	57.64	25.38	40.21	0
S12-DINO	81.72	49.37	36.18	88.61	50.56	51.15	34.81	48.66	56.47	25.62	41.23	1
S12-MAE	81.91	49.90	32.03	87.79	50.44	51.92	34.08	45.8	57.13	24.69	41.07	0
S12-Data2Vec	81.91	44.36	34.32	88.15	51.36	48.82	35.90	**54.03**	58.23	24.23	41.91	1
UNet Baseline	**84.51**	54.79	31.60	**91.42**	58.68	60.47	**39.46**	47.57	62.09	**46.34**	**35.67**	6
ViT Baseline	81.58	48.19	**38.53**	87.66	57.43	59.32	36.83	44.08	52.57	38.37	38.55	2

- UNet achieves strong results

Data scarcity

50% of labels

Model	HLS Burns	MADOS	PASTIS	Sen1Floods11	xView2	FBP	DynEarthNet	CropMap	SN7	AI4Farms	BioMassters	# Top-2
CROMA	81.52	57.68	32.33	90.57	51.44	48.01	38.30	42.20	59.31	28.19	38.50	4
DOFA	78.02	55.21	28.60	88.39	58.91	36.90	**39.20**	30.93	47.06	26.69	42.81	2
GFM-Swin	74.36	**63.37**	20.41	71.61	57.81	63.14	31.25	31.42	59.83	28.43	48.19	2
Prithvi	80.89	40.79	33.13	89.69	45.79	40.27	33.43	**42.51**	49.45	29.27	41.03	1
RemoteCLIP	74.28	53.26	17.46	71.67	57.43	**65.92**	30.91	36.3	50.83	25.11	50.09	1
SatlasNet	75.97	52.24	16.78	89.45	50.74	46.04	36.34	35.29	60.74	27.08	42.23	1
Scale-MAE	75.47	46.87	23.26	72.54	**59.45**	62.11	32.60	20.32	**61.24**	26.40	46.74	2
SpectralGPT	76.40	58.00	34.61	87.52	45.94	21.71	36.52	32.09	56.28	27.46	37.34	2
S12-MoCo	79.79	42.90	32.59	89.22	49.66	46.92	34.45	41.32	56.21	28.38	41.08	0
S12-DINO	80.12	40.42	**35.71**	88.93	48.46	44.85	32.76	31.13	55.14	25.68	41.47	1
S12-MAE	80.13	44.29	31.15	88.43	47.09	45.63	33.29	28.07	55.55	27.50	41.66	0
S12-Data2Vec	79.82	41.22	33.42	86.58	48.84	46.73	32.61	28.53	56.94	25.84	42.82	0
UNet Baseline	**82.39**	43.87	30.25	**90.91**	56.58	55.42	35.14	36.30	46.82	**45.02**	**36.72**	4
ViT Baseline	78.17	28.77	**38.71**	86.08	54.82	57.32	37.33	39.53	49.21	38.37	39.56	2

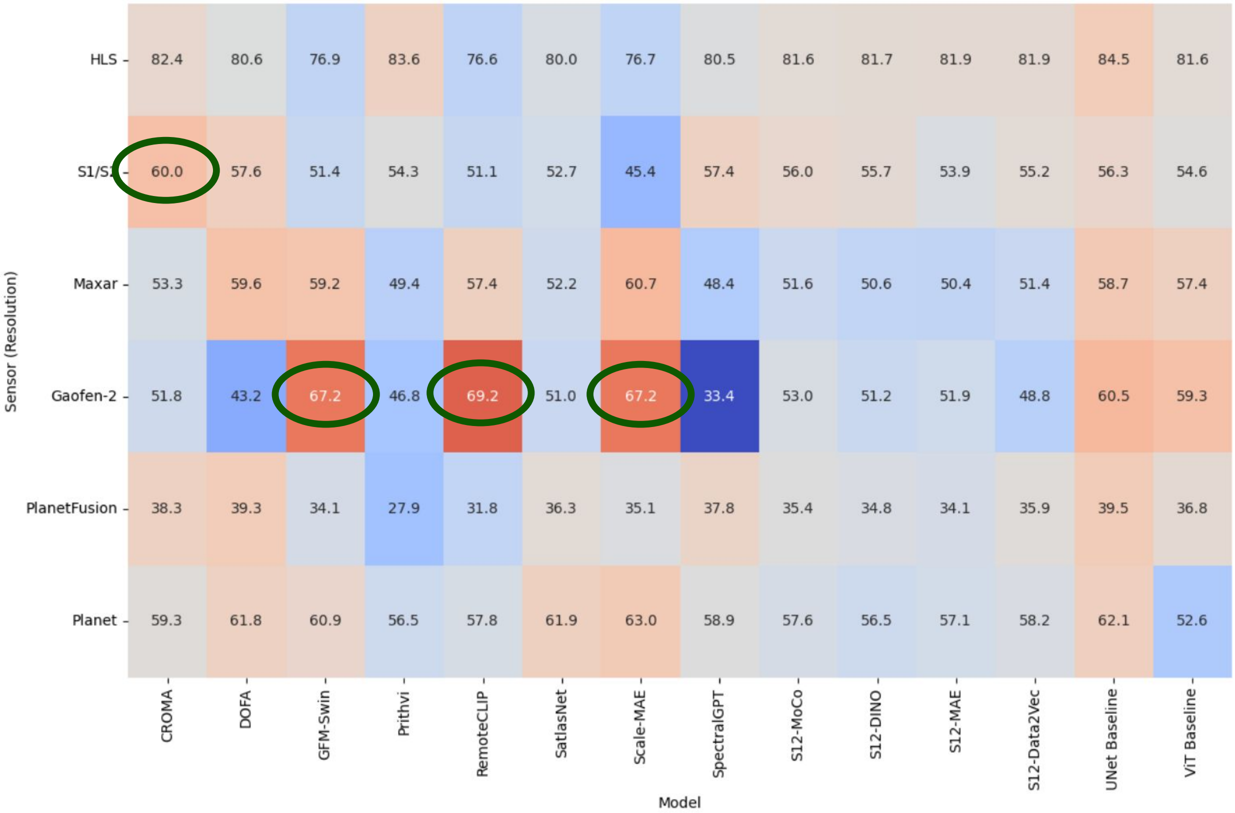
Data scarcity

10% of labels

Model	HLS Burns	MADOS	PASTIS	Sen1Floods11	xView2	FBP	DynEarthNet	CropMap	SN7	AI4Farms	BioMassters	#Top2
CROMA	76.44	**32.44**	32.80	87.22	46.54	37.39	36.08	36.77	42.15	38.48	40.25	6
DOFA	71.98	23.77	27.68	82.84	55.60	27.82	**39.15**	29.91	46.10	27.74	46.03	2
GFM-Swin	67.23	28.19	21.47	62.57	53.45	55.58	28.16	27.21	39.48	32.88	49.30	0
Prithvi	77.73	21.24	33.56	86.28	35.08	29.98	32.28	27.71	36.78	35.04	41.19	0
RemoteCLIP	69.40	20.57	17.19	62.22	53.75	56.23	34.43	19.86	43.11	23.85	53.32	1
SatlasNet	74.79	29.87	16.76	83.92	44.07	37.86	34.64	29.08	**49.78**	13.91	44.38	2
Scale-MAE	75.47	21.47	22.86	64.74	**56.06**	48.75	35.27	13.44	49.68	26.66	54.16	2
SpectralGPT	**83.35**	20.29	34.53	83.12	35.81	39.51	35.33	31.06	36.31	37.35	**39.44**	2
S12-MoCo	73.11	19.47	32.51	79.58	41.15	35.57	32.24	36.54	49.46	37.97	44.83	0
S12-DINO	75.93	23.47	36.62	84.95	41.02	34.63	32.78	**38.44**	41.15	37.91	42.74	2
S12-MAE	76.60	18.44	31.06	84.81	39.84	35.56	30.59	35.29	40.51	23.60	43.76	0
S12-Data2Vec	74.38	17.86	33.09	81.91	41.60	37.27	33.63	34.11	40.66	22.85	46.52	0
UNet Baseline	79.46	24.30	29.53	**88.55**	46.77	52.58	35.59	13.88	46.08	34.84	40.39	2
ViT Baseline	75.92	10.18	**38.44**	81.85	44.85	**56.53**	35.39	27.76	36.01	**39.20**	44.89	3

- UNet's performances drop in data-scarce scenarios
- Representations learned by GFMs are useful

Pretraining resolution



1. High-resolution data at pretraining is required to perform well on high-res. data at inference

2. Pretraining bands should match inference ones (not sensor agnostic)

PLAN

Part.1. PANGAEA

- Geospatial Foundation Models
- PANGAEA framework
- Evaluation Protocol
- Results
- Conclusion

Part.2. FlowEO

- Domain adaptation
- Flow Matching
- FlowEO
- Experiments & Results
- Conclusion

Conclusion

PANGAEA is an open-source codebase

README | Contributing | GPL-3.0 license

Tests passing

PANGAEA

paper arXiv | Pangaea Datasets | Pangaea Models | Pangaea Leaderboard

PANGAEA: A Global and Inclusive Benchmark for Geospatial Foundation Models

News

- [04/06/2025] We integrate Geo-Bench Datasets, including six segmentation and six classification tasks.
- [22/04/2025] on EarthDay, PANGAEA was officialy adopted to benchmark TerraMind. Read the news and the pre-print. We will release the benchmarking code in PANGAEA very soon!
- [05/12/2024] the pre-print is out!

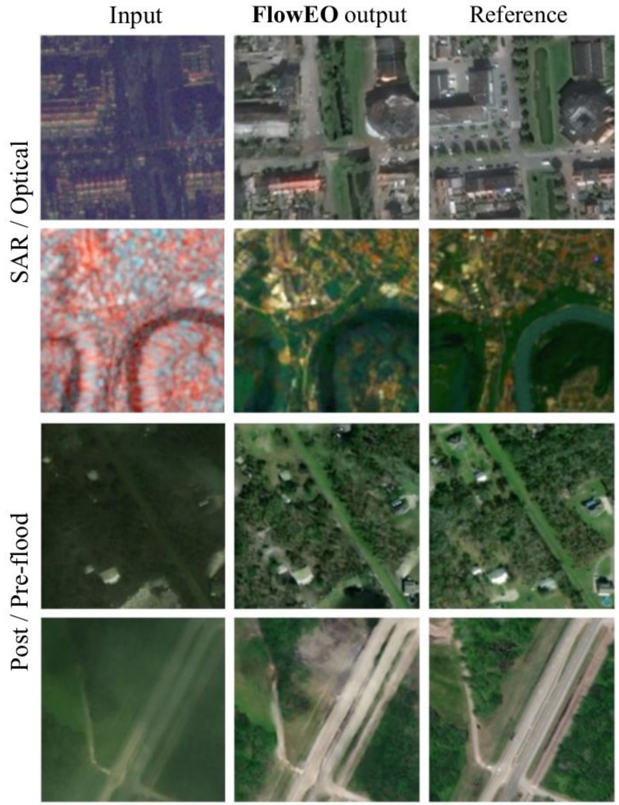
Includes:
- Datasets
- Models
- Decoder training
- Evaluation

They use PANGAEA:
- AnySat, G.Astruc et al. [CVPR2025]
- TerraMind, J.Jakubik et al. [ICCV2025]

Part.2

FlowEO: Generative Domain Adaptation for Earth Observation

Georges Le Bellier, Nicolas Audebert



Introduction

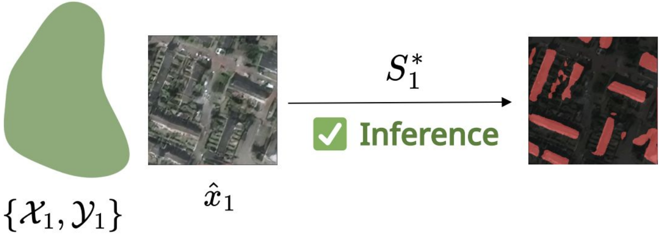
- Heterogeneous Earth observation data
- Distribution shifts

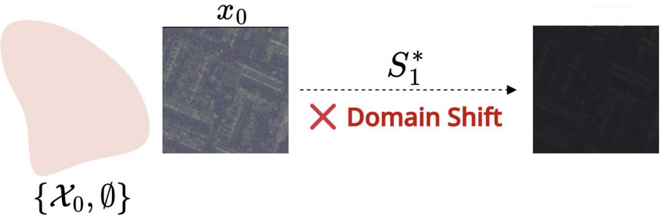
Sensor changes

Flooding

Different areas

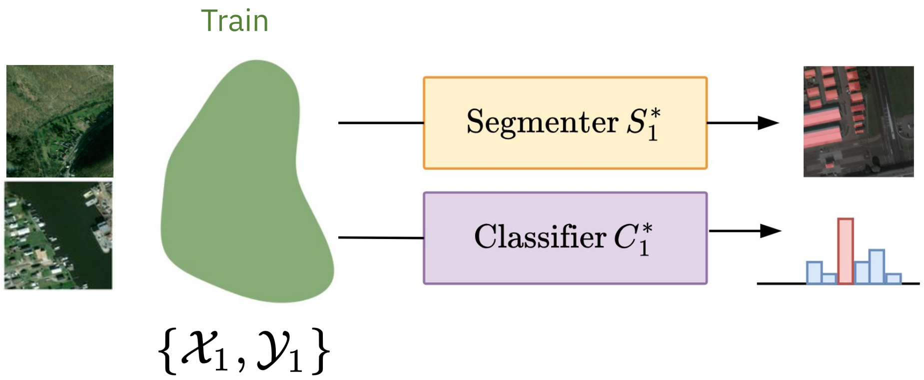
- Obstacle to pre-trained models inference





Introduction

- Frozen pre-trained predictive model / on-the-shelf model



We want a domain adaptation method that is

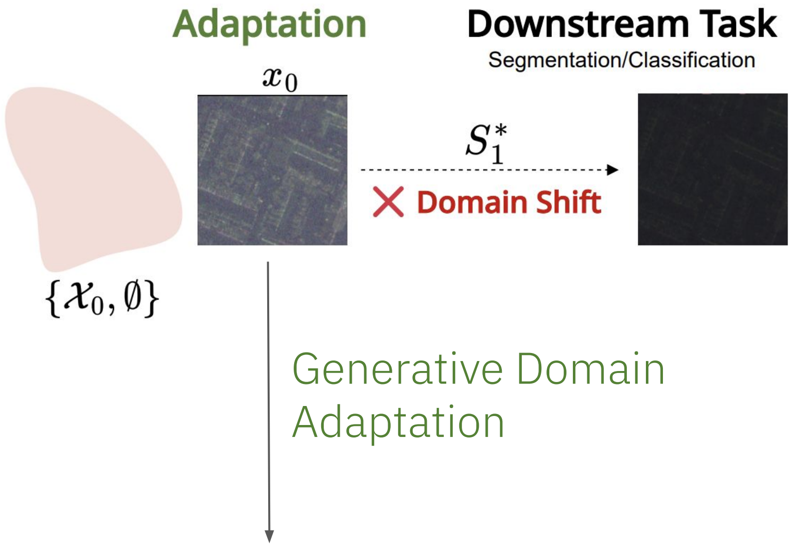
- Independent of the pre-trained model (architecture/features)
- Unsupervised (no label used)

PLAN

Part.1. PANGAEA
- Geospatial Foundation Models
- PANGAEA framework
- Evaluation Protocol
- Results
- Conclusion

Part.2. FlowEO
- Domain adaptation
- Flow Matching
- FlowEO
- Experiments & Results
- Conclusion

Generative Domain Adaptation



Pixel space adaptation
+ visual control
+ explainability
+ dense tasks

Main goal
→ Preserve semantic information

Idea
→ Leverage new generative models

PLAN

Part.1. PANGAEA

- Geospatial Foundation Models
- PANGAEA framework
- Evaluation Protocol
- Results
- Conclusion

Part.2. FlowEO

- Domain adaptation
- Flow Matching
- FlowEO
- Experiments & Results
- Conclusion

Flow Matching

- Generalization of diffusion models introduced in 2022 [2, 3, 4]
- Bridge arbitrary distributions p_0 and p_1 by learning a velocity field $u_t(\cdot)$

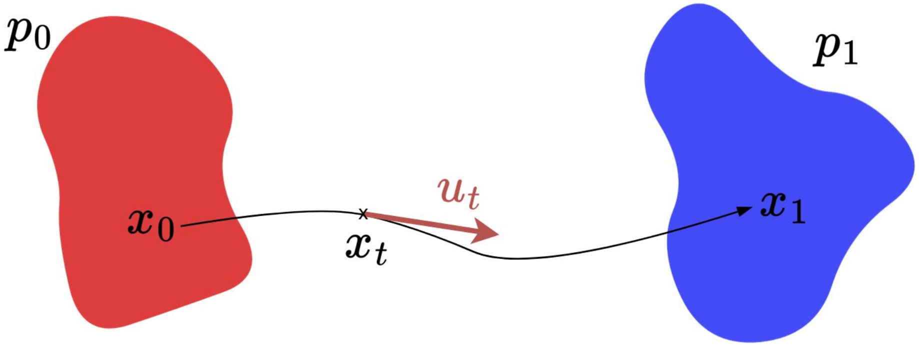
$$\frac{d}{dt}\varphi_t(x) = u_t(\varphi_t(x))$$

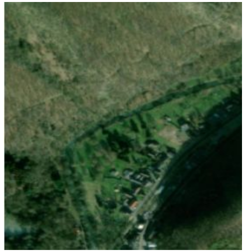
[2] Flow Matching for Generative Modeling, Y.Lipman et al.
[3] Building Normalizing Flows with Stochastic Interpolants, M.Albergo et al.
[4] Non-Denoising Forward-Time Diffusions, S.Peluchetti

Flow Matching

- Generalization of diffusion models introduced in 2022 [2, 3, 4]
- Bridge arbitrary distributions p_0 and p_1 by learning a velocity field $u_t(\cdot)$

$$\frac{d}{dt}\varphi_t(x) = u_t(\varphi_t(x))$$





$$x_1 = \varphi_1(x_0) = \mathrm{ODE}_u(x_0, t : 0 \to 1)$$

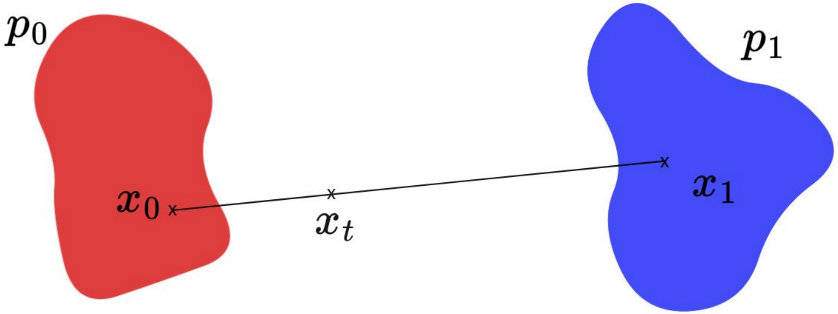
[2] Flow Matching for Generative Modeling, Y.Lipman et al.
[3] Building Normalizing Flows with Stochastic Interpolants, M.Albergo et al.
[4] Non-Denoising Forward-Time Diffusions, S.Peluchetti

Flow Matching

1. Sample $(x_0, x_1) \sim p(x_0, x_1)$

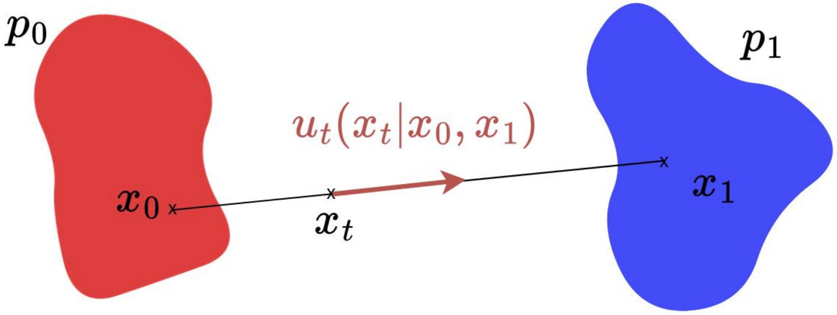
Flow Matching

1. Sample $(x_0, x_1) \sim p(x_0, x_1)$
2. Sample time $t \sim U(0, 1)$
3. Interpolant $x_t = (1 - t)x_0 + tx_1$



Flow Matching

1. Sample $(x_0, x_1) \sim p(x_0, x_1)$
2. Sample time $t \sim U(0, 1)$
3. Interpolant $x_t = (1 - t)x_0 + tx_1$
4. Simple regression loss



$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}\|v_\theta(t, x_t) - (x_1 - x_0)\|^2$$

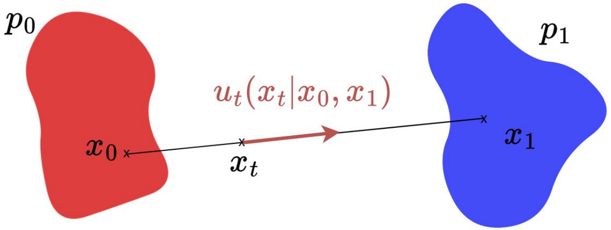
PLAN

Part.1. PANGAEA
- Geospatial Foundation Models
- PANGAEA framework
- Evaluation Protocol
- Results
- Conclusion

Part.2. FlowEO
- Domain adaptation
- Flow Matching
- FlowEO
- Experiments & Results
- Conclusion

Coupling and alignment

Choice of the coupling?



1. Sample $(x_0, x_1) \sim p(x_0, x_1)$

Semantic alignment

Unaligned

→ independent coupling

$p(x_0, x_1) = p(x_1)p(x_0)$

Weakly aligned

→ data dependent

$p(x_0, x_1) = p(x_1|x_0)p(x_0)$

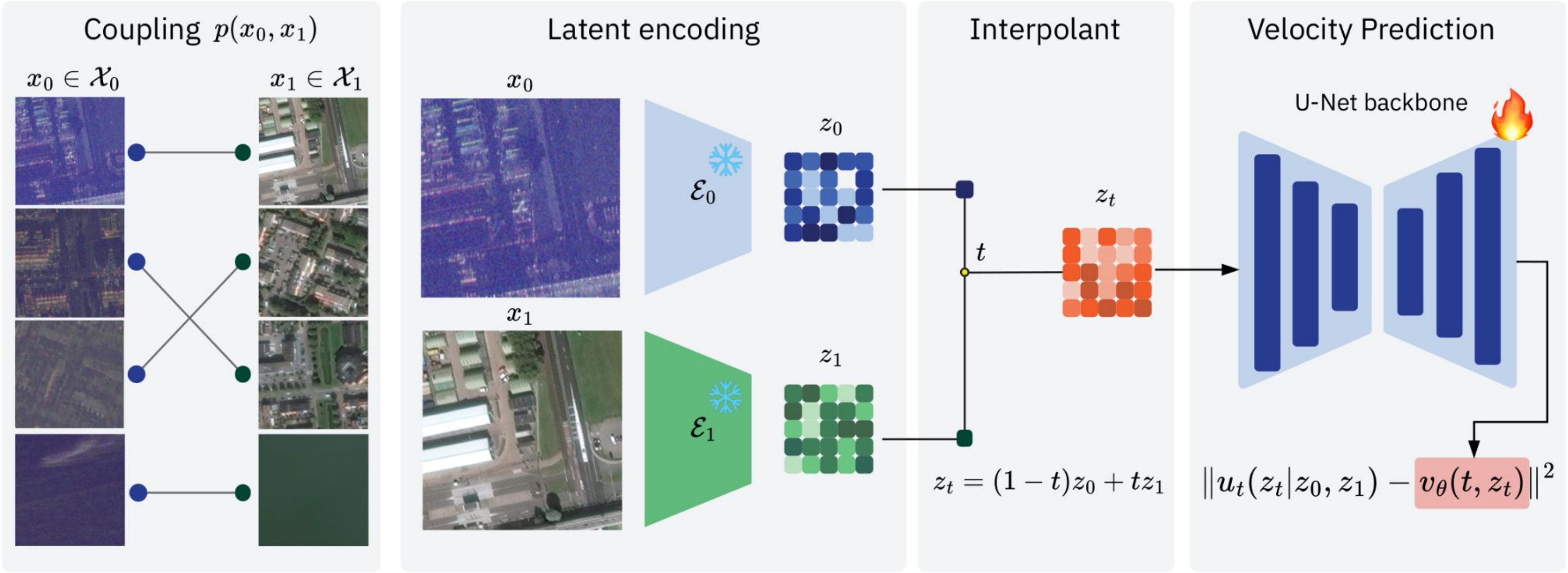
Strongly aligned

→ data dependent

$p(x_0, x_1) = p(x_1|x_0)p(x_0)$

FlowEO

Training

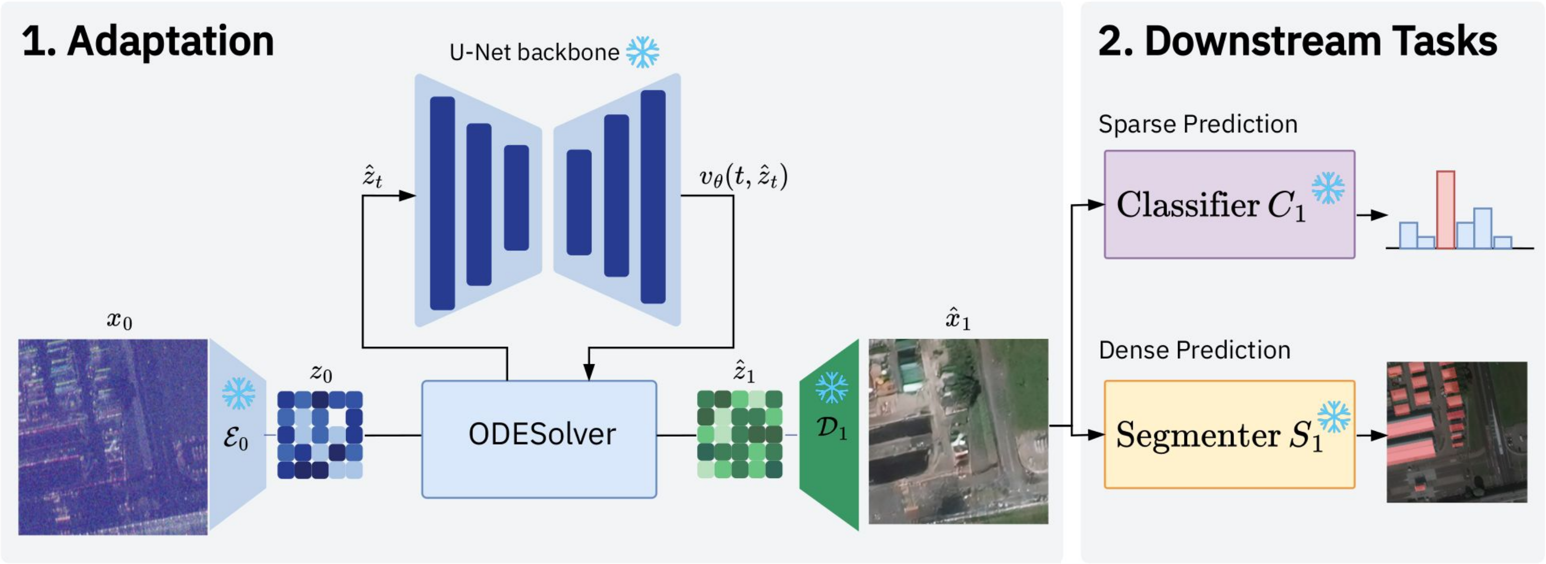


→ no predictive models, no labels used during training

FlowEO

Inference

Solve $\frac{d}{dt}\varphi_t(x) = u_t(\varphi_t(x))$ with $\boldsymbol{u}_t(\cdot)$ approximated by the learned model



PLAN

Part.1. PANGAEA

- Geospatial Foundation Models
- PANGAEA framework
- Evaluation Protocol
- Results
- Conclusion

Part.2. FlowEO

- Domain adaptation
- Flow Matching
- FlowEO
- Experiments & Results
- Conclusion

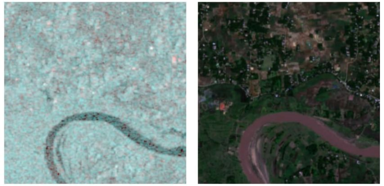
Experiments

Dataset	Target	Source	Resolution	Task	Size	Alignment
SpaceNet 6 [48]	SAR (aerial)	RGB (WorldView-2)	2 m/px	Segmentation	50 000	Strong
Sen1floods11 [4]	SAR (Sentinel-1)	Optical (Sentinel-2)	10 m/px	Segmentation	64 512	Strong
BigEarthNet2 (reBEN) [9]	SAR (Sentinel-1)	Optical (Sentinel-2)	10 m/px	Multi-label classification	237 871	Strong
SpaceNet 8 Germany [19]	RGB (post-flood)	RGB (pre-flood)	0.8 m/px	Segmentation	5688	Weak
SpaceNet 8 Louisiana [19]	RGB (post-flood)	RGB (pre-flood)	0.8 m/px	Segmentation	17 173	Weak

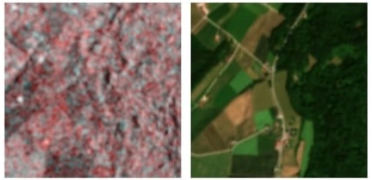
Strongly aligned:

SpaceNet 6

Sen1floods11



BigEarthNet2



Weakly aligned:

SpaceNet 8

Germany

Louisiana

Experiments

Baselines:

- Trained with data-dependent coupling
- Pix2pix, CycleGAN, StegoGAN, UNSB

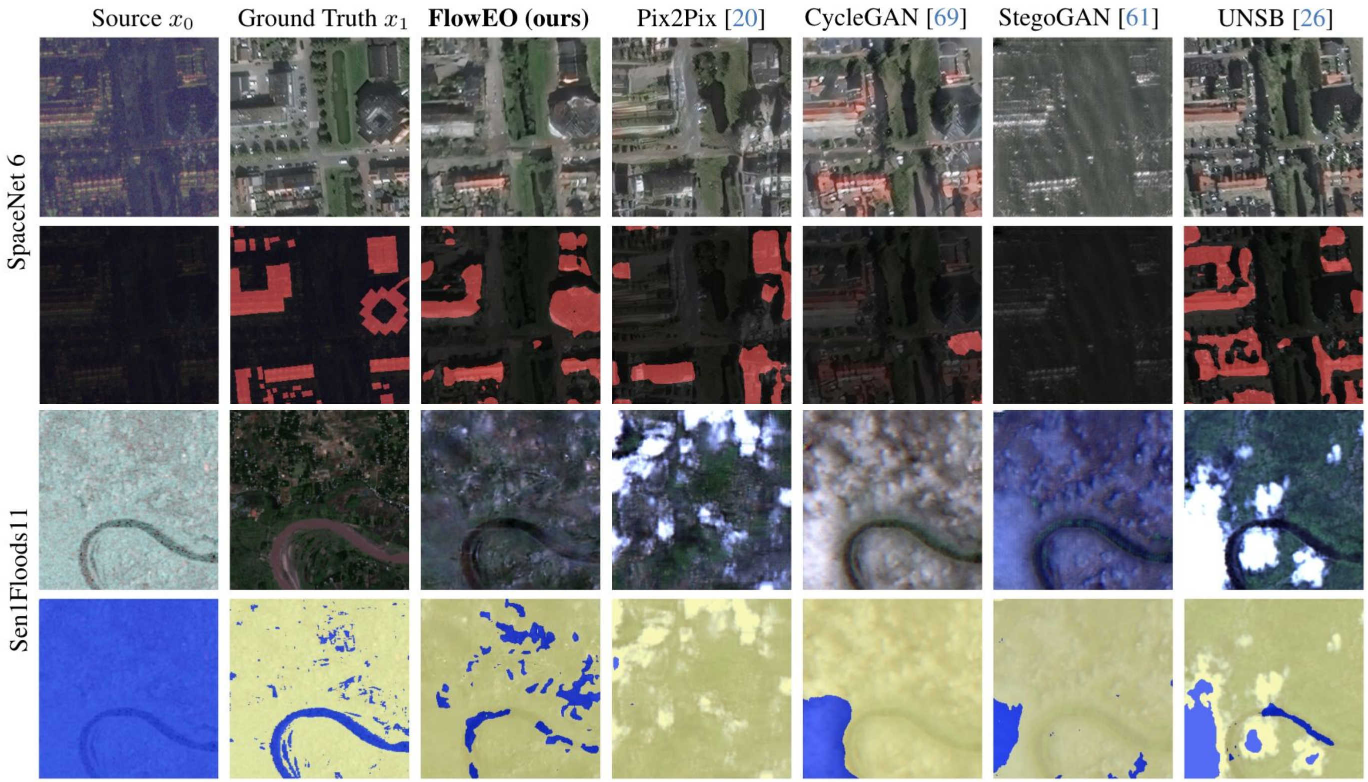
FlowEO:

- SD3 latent space (4, 32, 32)
- UNet backbone (same as for diffusion models)

Evaluation:

- Generation quality: LPIPS and FID
- Semantic preservation: mIoU, mAcc, F1

Results - strongly aligned

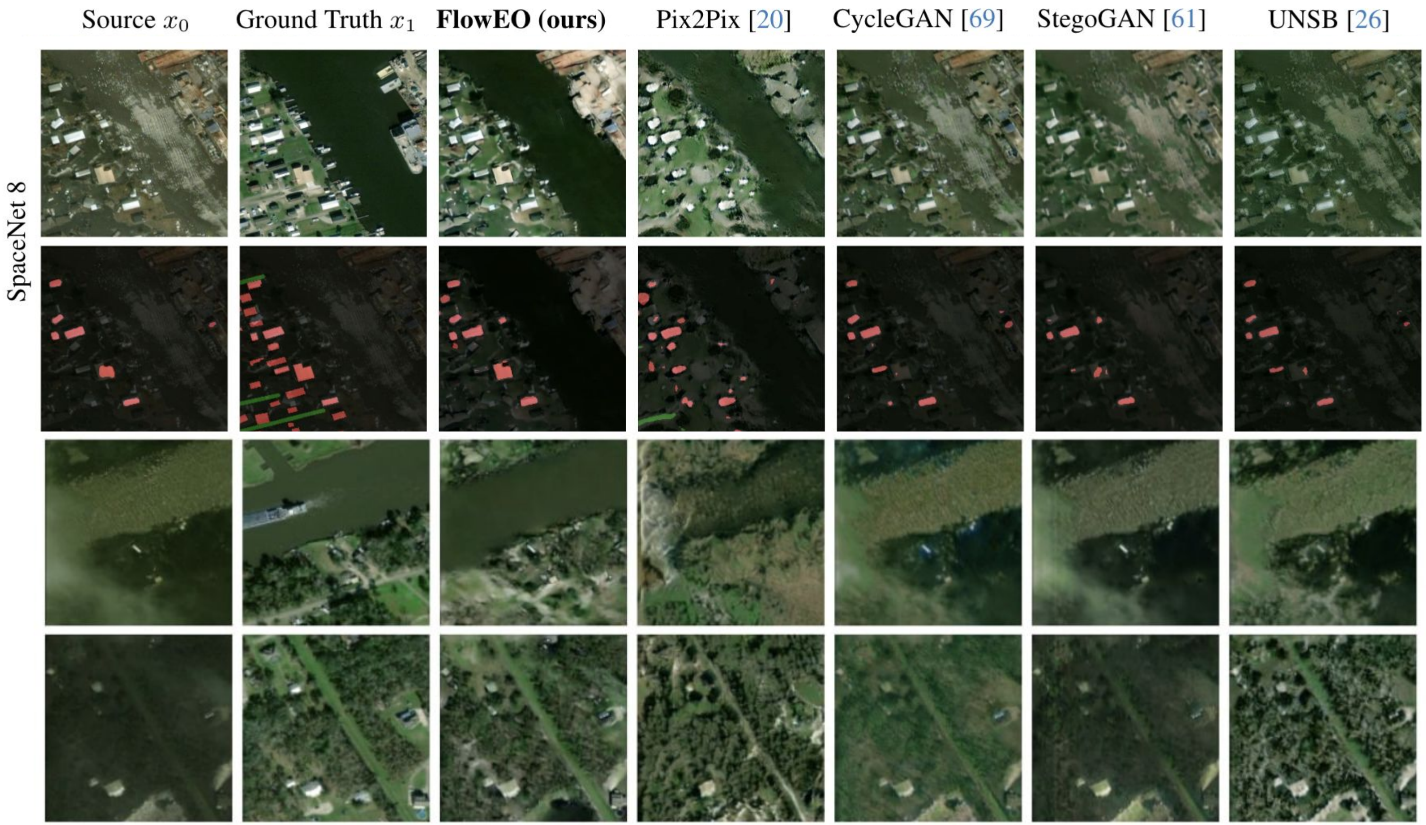


Results - strongly aligned

Datasets	Sen1Floods1				SpaceNet 6				ReBEN					
	SAR → Optical				SAR → RGB				SAR → Optical					
	mIoU	mAcc	FID	LPIPS	mIoU	mAcc	FID	LPIPS	AP^{μ}	AP^{M}	$F1^{\mu}$	$F1^{M}$	FID	LPIPS
No adaptation	06.22	49.72	297.22	84.84	31.94	41.01	275.05	79.48	17.46	17.43	02.31	01.31	339.36	85.99
Upper bound	55.14	71.28	00.00	00.00	84.94	90.74	00.00	00.00	79.26	65.28	74.28	62.84	00.00	00.00
Pix2Pix	51.50	62.31	20.64	31.33	56.48	63.43	130.42	41.89	**41.09**	27.88	**43.93**	**25.79**	**62.84**	17.56
CycleGAN	42.12	48.47	20.97	36.35	50.01	55.85	132.75	50.72	26.09	19.79	26.93	15.75	81.54	19.67
UNSB	42.69	48.85	23.01	35.01	52.43	61.04	**72.48**	45.81	25.61	20.71	29.52	19.45	113.73	35.64
StegoGAN	43.37	49.75	41.06	31.87	44.87	50.02	306.50	56.62	26.13	22.16	29.49	20.28	81.15	22.32
FlowEO	**54.92**	**69.04**	**12.96**	**29.21**	**65.07**	**72.33**	94.02	**39.96**	37.16	**32.14**	36.04	25.72	75.80	**15.51**

Table 3. Quantitative results on domain adaptation for strongly aligned datasets. We report both segmentation (mIoU, mAcc) or classification (AP/F1) and image quality metrics (FID, LPIPS). FlowEO preserves achieves the best UDA segmentation performances, and on-par classification performances with Pix2Pix.

Results - weakly aligned



Results - weakly aligned

Datasets	SpaceNet 8 Post-flood → Pre-flood				SpaceNet 8 Germany Post-flood → Pre-flood				SpaceNet 8 Louisiana Post-flood → Pre-flood			
	mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓	mIoU ↑	Acc ↑	FID ↓	LPIPS ↓	mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓
No adaptation	40.05	42.40	75.62	63.66	37.09	39.08	89.54	63.27	36.51	38.85	96.60	63.80
Upper bound	63.10	72.09	00.00	00.00	55.27	66.77	00.00	00.00	66.91	75.97	00.00	00.00
Pix2Pix	34.73	36.08	98.22	50.95	32.92	34.25	98.38	55.75	38.79	40.86	92.23	47.05
CycleGAN	40.70	43.35	**54.31**	55.70	39.35	41.79	**62.80**	59.46	42.39	45.14	**52.80**	52.92
UNSB	39.35	42.67	68.30	55.35	38.25	40.62	66.62	56.84	40.67	43.87	73.72	53.04
StegoGAN	38.62	40.58	66.61	58.07	36.74	38.78	90.42	63.50	40.14	42.29	68.56	54.58
FlowEO	**44.65**	**48.79**	60.32	**45.50**	**41.27**	**45.29**	82.74	**53.63**	**47.19**	**52.30**	59.65	**41.95**

Table 2. Quantitative results on domain adaptation for weakly aligned datasets. We report both segmentation (mIoU, mAcc) and image quality metrics (FID, LPIPS) for SpaceNet 8 and its geographic subsets. FlowEO transports images while preserving its semantics, achieving significant segmentation performance improvements in domain adaptation setting: 44.65 *vs*. 40.05 mIoU on SpaceNet 8. It also outperforms the second-best model – CycleGAN – on segmentation accuracy after transfer.

PLAN

Part.1. PANGAEA

- Geospatial Foundation Models
- PANGAEA framework
- Evaluation Protocol
- Results
- Conclusion

Part.2. FlowEO

- Domain adaptation
- Flow Matching
- FlowEO
- Experiments & Results
- Conclusion

Conclusion

- Flow matching improves segmentation scores.
- Works well for (weakly/strongly) aligned datasets
- FM models fail on non-aligned datasets to preserve semantic information.

Future Work

- New couplings for non-aligned datasets
- Dedicated VAE training
- New setups: data-augmentation/missing modality

Bonus: Part.1+2

FlowEO and GFMs: how to combine them ?

Missing modality: use generative models to generate the missing modality (e.g. in timeseries) to take advantage of multimodal EO models.

→ Similar to IBM's Thinking in Modality (TiM).

Thank for your attention

Results - impact of the coupling

SpaceNet 8				
	Post-flood $\rightarrow$ Pre-flood			
Coupling	mIoU $\uparrow$	mAcc $\uparrow$	FID$\downarrow$	LPIPS $\downarrow$
Independent $p(x_0, x_1)$	35.59	37.41	94.23	66.62
Minibatch-OT $\pi(x_0, x_1)$	37.26	39.28	84.44	63.93
Data-dependent $p(x_1\|x_0)p(x_0)$	**44.65**	**48.79**	**60.32**	**45.50**
	Pre-flood/Post-flood			
Coupling	mIoU $\uparrow$	mAcc $\uparrow$	FID $\downarrow$	LPIPS $\downarrow$
Independent $p(x_0, x_1)$	35.60	37.80	80.26	67.88
Minibatch-OT $\pi(x_0, x_1)$	36.21	39.28	73.26	65.22
Data-dependent $p(x_1\|x_0)p(x_0)$	**44.87**	**53.76**	**50.88**	**52.81**
SpaceNet 6				
	SAR $\rightarrow$ RGB			
Coupling	mIoU $\uparrow$	mAcc $\uparrow$	FID$\downarrow$	LPIPS $\downarrow$
Independent $p(x_0, x_1)$	45.25	50.75	145.02	65.94
Minibatch-OT $\pi(x_0, x_1)$	48.48	55.03	125.82	58.34
Data-dependent $p(x_1\|x_0)p(x_0)$	**65.07**	**72.33**	**94.02**	**39.98**
	RGB $\rightarrow$ SAR			
Coupling	mIoU $\uparrow$	mAcc $\uparrow$	FID $\downarrow$	LPIPS $\downarrow$
Independent $p(x_0, x_1)$	45.74	50.85	105.47	64.69
Minibatch-OT $\pi(x_0, x_1)$	47.25	52.65	91.74	60.24
Data-dependent $p(x_1\|x_0)p(x_0)$	**55.36**	**61.53**	**36.86**	**51.66**

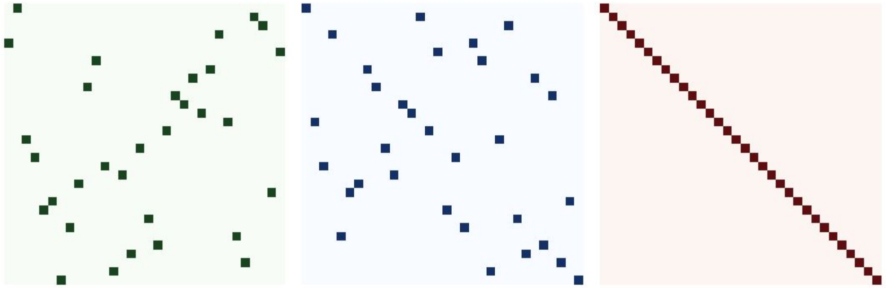


Figure 6. Comparison between the pairing matrices generated with the different couplings for a batch on SpaceNet 8, from left to right: independent coupling $p(x_0)p(x_1)$, OT-coupling $\pi(x_0, x_1)$, data-dependent coupling $p(x_1 \mid x_0)p(x_0)$.

Results - impact of VAE finetuning

		mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓
SpaceNet 8 Post-flood → Pre-flood					
RGB	Base	**44.65**	**48.79**	**60.32**	**45.50**
RGB	Finetuned	44.33	48.71	81.75	51.64
SpaceNet 6 SAR → RGB		mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓
RGB	Base	**65.07**	**72.33**	**94.02**	**39.96**
RGB	Finetuned	64.63	72.17	111.66	42.77
Sen1Floods11 SAR → Optical		mIoU ↑	mAcc ↑	FID ↓	LPIPS ↓
S2	Base	51.45	57.63	24.33	29.22
S2	Finetuned	**54.92**	**69.04**	**12.96**	**29.21**
ReBEN SAR → Optical		AP^M	$F1^M$	FID↓	LPIPS ↓
S2	Base	27.02	15.97	168.85	16.88
S2	Finetuned	**32.14**	**25.72**	**75.80**	**15.51**

Results - sampling

Sen1Floods11 SAR → Optical				
	mIoU ↑	mAcc ↑	FID↓	LPIPS ↓
25 Sampling Steps				
Linear	54.60	72.22	**13.99**	**28.91**
Sigmoid $\kappa = 10$	**55.05**	**72.50**	14.38	29.02
50 Sampling Steps				
Linear	54.26	71.79	**13.06**	**28.86**
Sigmoid $\kappa = 10$	**54.46**	**71.94**	13.46	28.90
100 Sampling Steps				
Linear	54.10	71.59	**12.87**	**28.85**
Sigmoid $\kappa = 10$	**54.19**	**71.66**	12.95	28.86
SpaceNet 6 SAR → RGB				
	mIoU ↑	mAcc ↑	FID↓	LPIPS ↓
25 Sampling Steps				
Linear	64.23	71.68	117.30	**42.78**
Sigmoid $\kappa = 10$	**64.46**	**71.93**	**113.64**	42.96
50 Sampling Steps				
Linear	63.98	71.46	119.68	**42.89**
Sigmoid $\kappa = 10$	**64.07**	**71.57**	**118.06**	42.98
100 Sampling Steps				
Linear	63.79	71.28	121.28	**42.98**
Sigmoid $\kappa = 10$	**63.83**	**71.34**	**120.38**	43.03