SkyCap: Bitemporal VHR Optical–SAR Quartets for Amplitude Change Detection and Foundation-Model Evaluation

Paul Weinmann **Ferdinand Schenck** **Martin Šiklar**
LiveEO GmbH
Berlin, Germany
{paul.weinmann, ferdinand.schenck, martin.siklar}@live-eo.com

Abstract

Change detection for linear infrastructure monitoring requires reliable high-resolution data and regular acquisition cadence. Optical very-high-resolution (VHR) imagery is interpretable and straightforward to label, but clouds break this cadence. Synthetic Aperture Radar (SAR) enables all-weather acquisitions, yet is difficult to annotate. We introduce **SkyCap**, a bitemporal VHR optical–SAR dataset constructed by archive matching and co-registration of (optical) **Sky**Sat and **Cap**ella Space (SAR) scenes. We utilize optical-to-SAR label transfer to obtain SAR amplitude change detection (ACD) labels without requiring SAR-expert annotations. We perform continued pretraining of SARATR-X on our SAR data and benchmark the resulting SAR-specific foundation models (FMs) together with SARATR-X against optical FMs on SkyCap under different preprocessing choices. Among evaluated models, MTP(ViT-B+RVSA), an optical FM, with dB+Z-score preprocessing attains the best result ($F1_c = 45.06$), outperforming SAR-specific FMs further pretrained directly on Capella data. We observe strong sensitivity to preprocessing alignment with pretraining statistics, and the ranking of optical models on optical change detection does not transfer one-to-one to SAR ACD. To our knowledge, this is the first evaluation of foundation models on VHR SAR ACD.

1 Introduction

Monitoring linear infrastructure such as pipelines, power lines, and railways requires reliable high-resolution change detection with consistent revisit intervals. For pipelines, this cadence is often mandated by regulation [12]. Due to its high interpretability, optical very high-resolution (VHR) imagery supports both reliable human labeling and effective machine learning analysis. However, it is fundamentally constrained by cloud cover, disrupting targeted revisit intervals [11].
Synthetic Aperture Radar (SAR), by contrast, enables imaging regardless of weather or daylight conditions, making it a robust alternative for consistent monitoring. To capture the small-scale changes relevant to infrastructure monitoring, we focus specifically on VHR SAR data. However, SAR interpretation poses significant challenges: complex backscatter patterns, speckle noise, and geometric distortions make direct annotation difficult, especially given limited expert capacity. These limitations motivate our exploration of SAR amplitude change detection (ACD) as a core component of infrastructure monitoring, aiming to combine SAR's acquisition reliability with foundation model (FM)-based CD methods.
Foundation models (FMs) have improved transfer learning in both vision [3] and remote sensing [2]. SAR-specific pretraining approaches such as SARATR-X adapt Masked Autoencoder pretraining to improve stability under speckle noise [9], but are evaluated primarily on few-shot object detection. It remains unclear how well such models transfer to the VHR SAR ACD task for infrastructure

monitoring, and how they compare to optical FMs applied to SAR with appropriate preprocessing. To address SAR labeling challenges, we construct *SkyCap*, a bitemporal optical–SAR dataset by cross-referencing Capella X-band Spotlight SAR with Planet SkySat optical pairs and transferring change labels from optical to co-registered SAR.

We study three questions: (1) How can we obtain reliable SAR change labels without SAR-expert annotation? (2) How do SAR pretrained foundation models compare to optical foundation models transferred to SAR for change detection? (3) Which preprocessing enables the best cross-modal transfer, and does the relative ranking of optical FMs translate to SAR?

Our contributions are: (i) a practical optical-to-SAR label transfer pipeline for VHR change detection; (ii) a controlled comparison of continued SAR pretraining on Capella and ALOS-2 SAR data against transferring optical FMs; (iii) an empirical result that optical FMs outperform SAR-specific pretraining on Capella Space Spotlight data, along with an analysis of preprocessing influence.

2 Related Work

SAR Amplitude Change Detection (ACD). Prior work on VHR amplitude-only SAR CD has focused on TerraSAR-X and COSMO-SkyMed under repeat-pass, matched-geometry pairs, with mostly qualitative case studies or limited metrics on small datasets [1, 7, 17]. **SAR Foundation Models.** SARATR-X [9], SAR-JEPA [8] and MSFA [10] explore improved representations for the Masked Autoencoder reconstruction target to improve the stability of MAE pretraining on speckle-afflicted SAR data. Multimodal approaches such as TerraMind [5] and DOFA [18] simultaneously pretrain on optical and SAR data. However, these multimodal approaches primarily target medium-resolution dual-polarization, C-band Sentinel-1 data rather than commercial VHR SAR (single-pol, X-band).

Optical Foundation Models for Remote Sensing. MTP [16] leverages multi-task pretraining on diverse annotated datasets, while USat [4] integrates multiple spectral bands and spatial resolutions. DINOv3 [14] introduces a general domain-agnostic training approach for optical data and the results on remote sensing data support the view that domain-agnostic pretraining can transfer effectively to specialized downstream domains [6]. Despite their success on optical data, systematic evaluation of their transfer to VHR SAR ACD remains unexplored.

3 Methodology

SkyCap Dataset Creation We build incidental, bitemporal optical–SAR quartets by archive-matching Capella Space Spotlight SAR (submeter GSD, X-band, HH polarization) and Planet SkySat optical scenes (0.5m GSD, RGB+NIR), then co-registering all scenes. We selected locations with a high probability of human-induced change, i.e., near human settlements and substantial temporal separation. For more details, see Appendix A. After location-based deduplication, we obtained 19 scene *quartets* (i.e. time step 1 and 2 each consist of a SkySat and Capella scene, in total four involved scenes) covering Eastern Europe, the Middle East and most of Asia across tropical, desert, and temperate biomes. This results in 3,484 annotated image-pair samples with changes. **Annotation Strategy** All change annotations were created on interpretable optical image-pairs by an experienced annotation team, then transferred to co-registered SAR pairs.

Continued Pretraining. We extend the SARATR-X pretraining pipeline by continuing pretraining on our SAR data using model weights obtained upon request from the authors, following their MSGF-based objective and configuration. SARATR-X is based on HiViT-B [19] with ImageNet initialization and modifies the MAE objective [3] by replacing backscatter intensities with Multi-Scale Gradient Features (MSGF). MSGF serve as a denoised representation of image content, reducing sensitivity to multiplicative speckle noise and stabilizing the pretraining objective. This yields three model variants trained for $\sim$73% of the SARATR-X training schedule: **CapellaX** was trained on 136k Capella X-band images. **ALOS-X** used 254k ALOS-2 L-band images. **CapALOS-X** combined both datasets in a 50/50 split. ALOS-X and CapALOS-X serve to evaluate cross-sensor transfer from ALOS-2/PALSAR-2 (10m; L-band) to Capella Space Spotlight (0.5m; X-band). We analyze how pretraining on these distinct SAR sources affects downstream performance on Capella data.

Change Detection Task We evaluate two tasks separately on SkyCap: a) Optical Change Detection on the SkySat pair from the SkyCap quartet utilizing the optical-derived labels, referred to as SkyCap optical b) SAR Amplitude Change Detection on the Capella pair from the same SkyCap quartets

Table 1: SkyCap SAR (Capella X-band). Change-class metrics (×100, higher is better). Best per FM underlined, best overall in bold.

Encoder	Preprocessing	$F1_c$	IoU_c	$Prec_c$	$Recall_c$
HiViT (optical)	linear	41.63	28.44	39.39	46.86
	linear+Z-score	41.71	28.60	38.63	48.41
	dB+Z-score	42.11	29.04	39.15	48.78
SARATR-X	linear	40.03	27.11	36.99	46.19
	dB+Z-score	34.97	23.14	32.04	42.22
CapellaX	linear	42.06	28.90	38.86	48.44
	dB+Z-score	38.76	26.15	32.04	48.95
CapALOS-X	linear	44.35	30.87	41.81	49.94
	dB+Z-score	39.70	27.04	36.53	46.13
MTP ViT+RVSA (optical)	linear	42.20	29.35	40.53	48.93
	linear+Z-score	44.52	31.12	40.60	52.22
	dB+Z-score	**45.06**	**31.68**	41.73	51.51
DINOv3 ViT (optical)	linear	41.60	28.80	38.80	48.63
	linear+Z-score	41.20	28.41	37.42	49.37
	dB+Z-score	42.40	29.18	36.84	**53.52**
DINOv3 ConvNeXt (optical)	linear	43.53	30.47	40.80	49.27
	linear+Z-score	44.25	31.03	41.92	49.29
	dB+Z-score	44.18	31.02	**45.57**	44.96

Table 2: Results on SkyCap optical (SkySat). Change-class metrics (×100, higher is better). Best overall in bold.

Encoder	F1	IoU	Prec.	Recall
HiViT (optical)	63.31	48.18	60.15	68.67
MTP(ViT+RVSA)	60.86	45.44	57.05	66.46
DINOv3 ViT	66.07	50.65	63.40	**69.96**
DINOv3 ConvNeXt	**68.18**	**53.25**	**67.25**	69.82

utilizing the same optical-derived labels as in a), referred to as SkyCap SAR. This allows us to directly compare SAR ACD with optical CD.

Change Detection Architecture. We employ a simple middle-fusion siamese architecture with the respective investigated encoder as the backbone, an absolute difference neck, and a U-Net [13] decoder. This matches the model architecture utilized for Change Detection in MTP [16].

SAR Preprocessing. We evaluate three preprocessing configurations: (1) *linear*: clip amplitudes to 0.5-99.5 percentiles of the training distribution and scale to [0,1], this aims to follow the preprocessing employed in SARATR-X; (2) *linear+Z-score* : linear preprocessing followed by Z-score normalization; (3) *dB+Z-score* : converts amplitudes to decibels, followed by applying Z-score normalization. The decibel scaling transforms the right-skewed gamma-distributed SAR intensities into an approximately normal distribution, more closely matching the distribution of natural optical images. SARATR-X uses linear preprocessing, and its multi-scale gradient features are not directly compatible with dB inputs, motivating our linear-input setting for SAR-pretrained models.

Training. We evaluate six encoders in total, three of which are optical, i.e. HiViT [19], MTP(ViT-B+RVSA)[16], and DINOv3 [14], and three SAR pretrained models SARATR-X [9], CapellaX, and CapALOS-X. CapellaX and CapALOS-X both follow the training approach of SARATR-X and were further pretrained on Capella Space sensor data. We limit the evaluation on optical data to optical models. All evaluated encoders are of the Base size ($\sim$90M parameters) for a fair comparison. For more details, see Appendix B.

4 Results

We report $\mathbf{F1_c}$, $\mathbf{IoU_c}$, $\mathbf{Prec_c}$ (precision), $\mathbf{Recall_c}$ for the *change* class only multiplied by 100. All absolute differences are expressed in percentage points (pp).

4.1 Change Detection Results

SkyCap (Capella X-band). Table 1 reports test results. With *linear* inputs, all optical models outperform SARATR-X (F1 = 40.03). Continued pretraining on Capella improves performance (CapellaX F1 = 42.06, +0.43 pp vs HiViT, +2.03 pp vs SARATR-X), and CapALOS-X is the strongest SAR-pretrained model (F1 = 44.35, +2.72 pp vs HiViT, +4.32 pp vs SARATR-X). MTP(ViT-B+RVSA) with *dB+Z-score* achieves the best overall result (F1 = 45.06). Preprocessing matters: *dB+Z-score* improves optical Transformers by +2.32 to +2.86 pp but reduces SAR-pretrained models by -1.27 to -5.06 pp. For DINOv3 ConvNeXt-B, *linear+Z-score* slightly exceeds *dB+Z-score*.

SkySat (optical) data. For comparison, we report results on the optical parts of the quartets from which the annotations were obtained in Table 2. The models were trained in a similar fashion to the SAR case, but no continued pretraining was performed. The best results were achieved by the ConvNeXt version of DINOv3 [14] with an F1 score of 68.18.

5 Discussion and Conclusion

Q1. We obtain reliable SAR change labels without SAR-expert annotation by transferring labels from interpretable optical imagery to co-registered SAR within our multimodal quartets, producing SAR data to train binary ACD models and systematically evaluate the impact of foundation-model backbones (encoders).
Q2. On high-resolution Capella X-band, optical foundation models with *dB+Z-score* achieve the best performance (best: MTP, F1 = 45.06), outperforming SAR-specific approaches including continued pretraining on target-sensor data. This is surprising as the SAR ACD data is in distribution for the SAR FMs and out of distribution for the optical FMs. We hypothesize that optical FMs outperform the evaluated SAR-specific models in our setting because dB+Z-score preprocessing reshapes the roughly gamma-distributed SAR amplitudes into an approximately normal distribution closer to natural images and enhances low-intensity structural contrast, whereas the SARATR-X-style pretraining reconstruction target (MSGF) emphasizes bright scatterers, which is helpful for Automatic Target Recognition but less well aligned with the subtle, low-backscatter changes that dominate SkyCap.
Q3. Preprocessing alignment is central. Models perform best when evaluation inputs match their pretraining statistics. Optical FMs gain under *dB+Z-score*, while SAR-pretrained models trained on linear amplitudes lose accuracy under dB inputs. The ranking from optical CD does not transfer one-to-one to SAR ACD: DINOv3 ConvNeXt leads on optical, MTP(ViT+RVSA) leads on SAR. We hypothesize that architectural factors, for example MTP's Rotated Variable Sized Attention [15], contribute more to transfer performance than differences in optical feature quality alone. The gap between the best optical result (F1 = 68.18) and the best SAR result ($F1_c$ = 45.06) is 23.12 pp, underscoring a persistent modality gap between VHR optical and SAR amplitude change detection. These findings indicate that careful input transformation can substitute for costly SAR-specific pretraining in VHR SAR amplitude change detection, and that cross-modal label transfer is a practical path to create evaluation data at scale.

5.1 Limitations

To our knowledge, this is the first evaluation of foundation models for very high resolution (VHR) SAR amplitude change detection (ACD). The findings should therefore be viewed as indicative rather than conclusive. The dataset's size and geographic coverage remain limited, preventing strong claims about statistical significance or global generalization. We rely on random patch-level splits instead of geographically disjoint splits, which constrains conclusions about spatial generalization. Despite manual refinement, small co-registration errors between optical and SAR imagery introduce residual geometric misalignments and label noise. Moreover, SAR backscatter often diverges from true object geometry, causing discrepancies between optical labels and SAR signal responses. Temporal offsets of up to five days between optical and SAR acquisitions may further introduce mismatched change labels, and not all optically visible changes necessarily yield measurable X-band backscatter differences.

5.2 Future Work

Key directions include: (1) Investigating why optical foundation models transfer successfully despite fundamental physical differences between optical and SAR sensing; (2) Further improving *SkyCap* by (a) adapting optical-derived label masks to better match SAR backscatter responses, and (b) analyzing which object classes or change types are poorly represented or invisible in SAR backscatter and refining the label set accordingly; (3) Exploring multimodal pretraining strategies that jointly learn from optical and SAR imagery to capture complementary information.

In conclusion, we demonstrate a path toward reliable all-weather change detection for infrastructure monitoring by combining optical supervision with SAR acquisition through multimodal dataset creation, and show that optical foundation models with appropriate preprocessing can outperform SAR-specific pretraining on very high-resolution data.

Acknowledgments

This work was supported by the German Federal Ministry for Economic Affairs and Climate Action (grant 50EE2016). We thank Capella Space and Planet Labs PBC for providing access to archival imagery. Any opinions and conclusions are those of the authors and do not necessarily reflect the views of the data providers.

References

[1] Markus Boldt and Karsten Schulz. Change detection in high resolution sar images: Amplitude based activity map compared with the covamcoh analysis. In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pages 3803–3806, 2012. doi: 10.1109/IGARSS.2012.6350584.

[2] Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David Lobell, and Stefano Ermon. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 197–211. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/01c561df365429f33fcd7a7faa44c985-Paper-Conference.pdf.

[3] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.

[4] Jeremy Irvin, Lucas Tao, Joanne Zhou, Yuntao Ma, Langston Nashold, Benjamin Liu, and Andrew Y Ng. Usat: A unified self-supervised encoder for multi-sensor satellite imagery. *arXiv preprint arXiv:2312.02199*, 2023.

[5] Johannes Jakubik, Felix Yang, Benedikt Blumenstiel, Erik Scheurer, Rocco Sedona, Stefano Maurogiovanni, Jente Bosmans, Nikolaos Dionelis, Valerio Marsocci, Niklas Kopp, et al. Terramind: Large-scale generative multimodality for earth observation. *arXiv preprint arXiv:2504.11171*, 2025.

[6] Saad Lahrichi, Zion Sheng, Shufan Xia, Kyle Bradbury, and Jordan Malof. Is self-supervised pre-training on satellite imagery better than imagenet? a systematic study with sentinel-2. *arXiv preprint arXiv:2502.10669*, 2025.

[7] Sofia Lanfri, Marcelo Scavuzzo, Mario A Lanfri, Gabriela Palacio, and Alejandro C Frery. Change detection methods in high resolution cosmo skymed images. In *Conference Proceedings of 2013 Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, pages 304–307. IEEE, 2013.

[8] Weijie Li, Wei Yang, Tianpeng Liu, Yuenan Hou, Yuxuan Li, Zhen Liu, Yongxiang Liu, and Li Liu. Predicting gradient is better: Exploring self-supervised learning for sar atr with a joint-embedding predictive architecture. *ISPRS Journal of Photogrammetry and Remote Sensing*, 218:326–338, 2024.

[9] Weijie Li, Wei Yang, Yuenan Hou, Li Liu, Yongxiang Liu, and Xiang Li. Saratr-x: Towards building a foundation model for sar target recognition. *IEEE Transactions on Image Processing*, 2025.

[10] Yuxuan Li, Xiang Li, Weijie Li, Qibin Hou, Li Liu, Ming-Ming Cheng, and Jian Yang. Sardet-100k: Towards open-source benchmark and toolkit for large-scale sar object detection. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

[11] Tatiana Nazarova, Pascal Martin, and Gregory Giuliani. Monitoring vegetation change in the presence of high cloud cover with sentinel-2 in a lowland tropical forest region in brazil. *Remote Sensing*, 12(11), 2020. ISSN 2072-4292. doi: 10.3390/rs12111829. URL https://www.mdpi.com/2072-4292/12/11/1829.

[12] Pipeline and Hazardous Materials Safety Administration (PHMSA). 49 cfr §192.705 – transmission lines: Patrolling; and 49 cfr §195.412 – inspection of rights-of-way and crossings under navigable waters. https://www.ecfr.gov/current/title-49, 2023. Code of Federal Regulations, Title 49, 2023 edition.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.

[14] Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. DINOv3, 2025. URL https://arxiv.org/abs/2508.10104.

[15] Di Wang, Qiming Zhang, Yufei Xu, Jing Zhang, Bo Du, Dacheng Tao, and Liangpei Zhang. Advancing plain vision transformer toward remote sensing foundation model. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023. doi: 10.1109/TGRS.2022.3222818.

[16] Di Wang, Jing Zhang, Minqiang Xu, Lin Liu, Dongsheng Wang, Erzhong Gao, Chengxi Han, Haonan Guo, Bo Du, Dacheng Tao, et al. Mtp: Advancing remote sensing foundation model via multitask pretraining. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:11632–11654, 2024.

[17] Diana Weihing, Felicitas von Poncét, Michael Schlund, and Oliver Lang. Change analysis with terrasar-x data. In W. Wagner and B. Székely, editors, *ISPRS TC VII Symposium – 100 Years ISPRS*, volume XXXVIII of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (IAPRS)*, pages 644–647, Vienna, Austria, July 2010. ISPRS. URL https://isprs.org/proceedings/XXXVIII/part7/b/pdf/644_XXXVIII-part7B.pdf.

[18] Zhitong Xiong, Yi Wang, Fahong Zhang, Adam J Stewart, Joëlle Hanna, Damian Borth, Ioannis Papoutsis, Bertrand Le Saux, Gustau Camps-Valls, and Xiao Xiang Zhu. Neural plasticity-inspired foundation model for observing the Earth crossing modalities. *arXiv preprint arXiv:2403.15356*, 2024.

[19] Xiaosong Zhang, Yunjie Tian, Lingxi Xie, Wei Huang, Qi Dai, Qixiang Ye, and Qi Tian. Hivit: A simpler and more efficient design of hierarchical vision transformer. In *International Conference on Learning Representations*, 2023.

A Dataset Details

Table 3: SAR Imaging Geometry Constraints

Parameter	**Requirement**
Orbit	Identical
Orbit direction	Identical
Observation direction	Identical
Look/squint angle difference	≤ 2 °

Table 4: Temporal Constraints

Interval Type	**Time Range**
Short-term	14-105 days
Medium-term	$\sim$1 year $\pm$ 45 days
Long-term	$\sim$2 years $\pm$ 45 days

Table 5: Optical Image Requirements

Parameter	**Requirement**
Temporal proximity to SAR	Within 5 days
Cloud coverage	$\geq 90\%$ cloud-free
Off-nadir angle	< 30 °
Intersection over Union (IoU) with SAR footprint	$\geq 60\%$

Table 6: SkyCap Dataset Composition

Metric	**Value**
Minimum overlap area	$\geq$ 15 km²
Number of quartets (after deduplication)	19
Geographic coverage	Eastern Europe, the Middle East, parts of Asia
Biomes covered	Tropical, desert, temperate
Co-registration accuracy (MAE)	5 pixels (2.5m)
Keypoints used for registration	30-50
Tile size	512×512 pixels
Total image-pairs (pretraining)	**68,000**
Annotated image-pair samples with changes (change detection)	**3,484**

B Training details

All models are fully end-to-end fine-tuned on images of size 512x512 pixels for 50 epochs with AdamW optimizer. The SAR models are trained with a learning rate of $1e-5$, weight decay of 0.05, batch size of 64, a cosine annealing scheduler with 20% linear warmup, and combined class-weighted Dice + weighted cross-entropy loss. The random initialized U-Net decoder receives an LR multiplier of 10. The optical models follow the same protocol, except for a batch size of 16, learning rate of $3e-5$ and weight decay of 0.1. We follow a standard 70%/10%/20% train/val/test split with random patch-level splits within each scene quartet, due to the small number and diversity of SkyCap quartets making a representative scene-level split infeasible. All models were trained on a single NVIDIA L40S with 48GB of VRAM, and training takes between 4.5 and 6.5 hours, depending on the backbone.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Our claim that we create a new dataset is explained in Section 3, while the results we claim to achieve are stated in Table 1

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations of our approach are candidly discussed in Section 5.1.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

Justification: This is an applied work and as such makes no novel theoretical claims.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: In Section 3(Methodology), as well as Appendices A and B, we describe the steps necessary to build our dataset as well as train our models. The foundation models our models are based on all have open weights, and the model architecture we use can be found in the source code for the relevant reference publication.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: The dataset we build is made up of satellite data which we are licensed to use, but not redistribute. Release of the dataset would require a release from the license by both relevant providers (Planet and Capella). As for the code: this is a work of applied ML and the code will not be released for reasons of IP.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: Hyperparameters are shared where applicable. Dataset splits are described, but as the dataset is not available, splits are only described in general, and no concrete splits can be released.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: Training each configuration is computationally intensive on our hardware, so we could not afford multiple runs per setting to estimate variance. We therefore do not report error bars.

Guidelines:

- The answer NA means that the paper does not include experiments.

- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Details of how the models were trained are provided in Appendix B.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Our research follows the code of ethics set by NeurIPS.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: Our work focuses on monitoring of energy infrastructure. To our understanding there is not a broader societal impact outside of possibly reducing the cost and improving the quality of monitoring this infrastructure.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: As the models are not released there are no guardrails needed. Beyond that, we foresee no risk of misuse even if that was the case.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: For the foundation models and architectures we use, the relevant publications are cited and the licenses adhered to. The data is not released largely due to licensing constraints from the data providers.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.

- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: No new assets are released along with this work.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: No crowdsourcing was used, and no human subjects were involved in this work.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: No human subjects were involved in this research.

    Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: No core development of this research involves LLMs.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.