Evaluating TerraMind for Conflict Damage Mapping Using Sentinel-2 Imagery

ML4EO 2026

23.06.2026

Clara Clipea, Dr. Daniel Racek

Universität der Bundeswehr München

Kompetenzzentrum Krisenfrüherkennung

The only scary thing today is...

... cloud masking

Objective & Motivation

- **Predict different levels of damage after conflict**
- Satellite imagery in Gaza
- Scalable, reproducible pipeline
- Freely available data

- Contribution:
  - **First use of an EO foundational model -** TerraMind
  - In conflict damage mapping

Context of Conflict Damage Research

Kompetenzzentrum Krisenfrüherkennung

- The analysis of infrastructure damage **has grown increasingly significant** (Sowers, Weinthal, & Zawahri, 2017, p. 410).

- Most existing studies rely on:
  - either **very high-resolution** (VHR) satellite imagery such as Maxar or Google Earth (Risso et al., 2024; Bouabid & Farah, 2024; Holali et al., 2024, Mueller et al., 2021)
  - or **medium-resolution** data such as Sentinel-2 (Gazel-Anthoine, 2022; Aimaiti et al., 2022, Pfeifle,2022)

Advantages of EO Foundation Models

Kompetenzzentrum Krisenfrüherkennung

Pre-trained on global data

Reduce task-specific training

Generalization

Efficiency

TerraMind

Jakubik et al., 2025

a large-scale, *any-to-any* **generative**, **multimodal** EO foundation model over nine modalities, it enables generation plus zero-/few-shot use

Zero-shot tasks: it tests representation quality without finetuning via (i) water-body mapping (pixel-wise) and (ii) geo-localization.

Few-shot classification: 1-shot / 5-shot image classification on EuroSAT & METER-ML using encoder embeddings.

Multiple Approaches

Training-free (few-shot)

Freeze encoder → patch level embeddings

Build tiny per-class prototypes from UNOSAT pixels

Predict by nearest prototype

Small supervised head

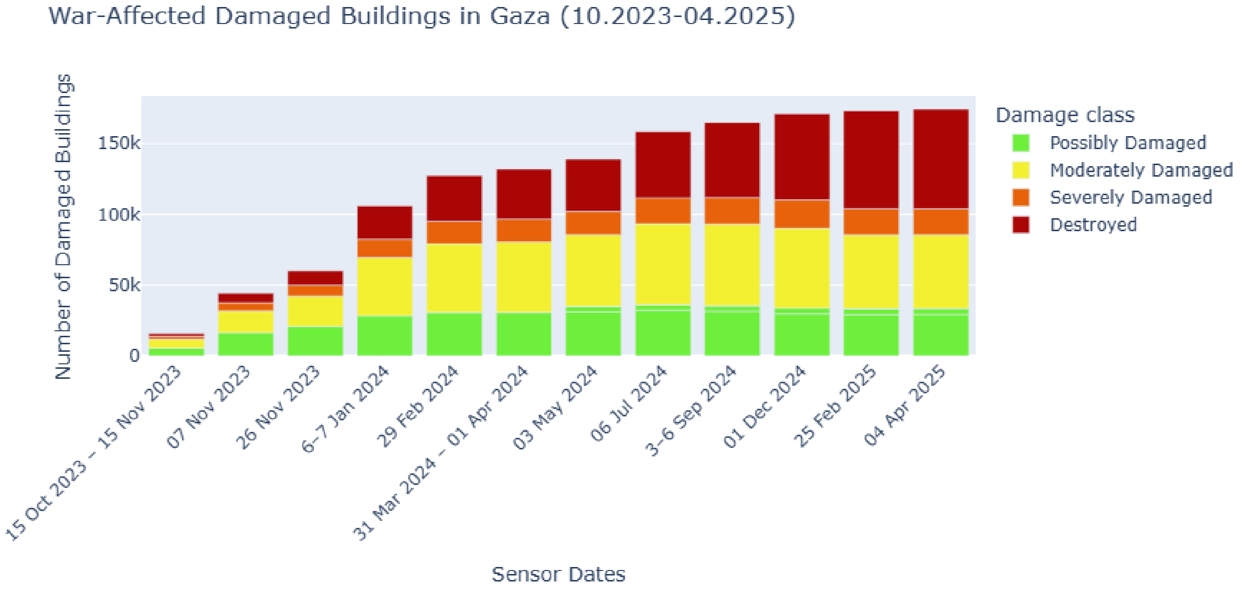
Freeze (or fine-tune) encoder, train a small classifier on embeddings

Uses UNOSAT labels, standard cross-entropy

More setup, better boundary shaping, but needs tuning

UNOSAT Labels

- 5 classes
  - No damage
  - Possibly
  - Moderate
  - Severe
  - Destroyed
- **No damage class -** needs to be defined on its own



No Damage Label

- Across prior work, “no-damage” is illustrated in several ways:
  - Pfeifle (2023)—and Aimaiti et al. (2022)— background within the building/AOI mask;
  - Scher & Van Den Hoek (2025) - implicit background where SAR coherence stays stable ≥3 months in built-up, non-mining areas;
  - Mueller et al. (2021) model only the destroyed class within populated urban areas, treating all confidently non-destroyed patches as negatives and masking temporally ambiguous patches via temporal/spatial smoothing

- **My approach**: define “no-damage” using pre-event pixels of buildings that are later labeled damaged, and “damage” using their post-event pixels.

Input masks

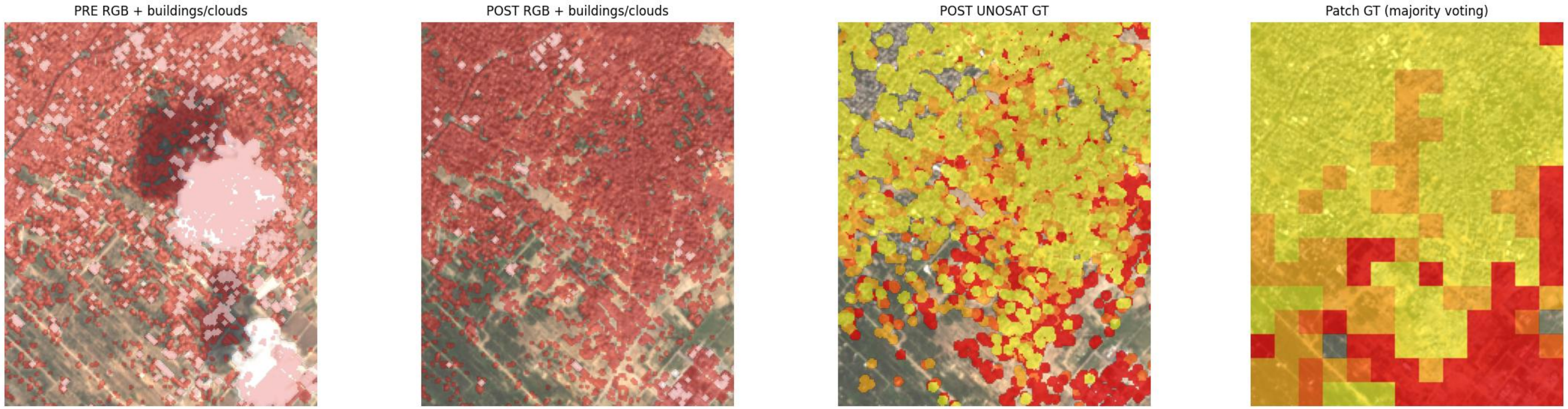
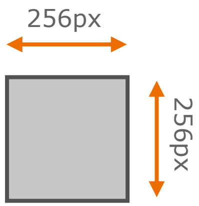

PRE RGB + buildings/clouds
POST RGB + buildings/clouds
POST UNOSAT GT
Patch GT (majority voting)

Image Preprocessing

Kompetenzzentrum
Krisenfrüherkennung

Data & AOI



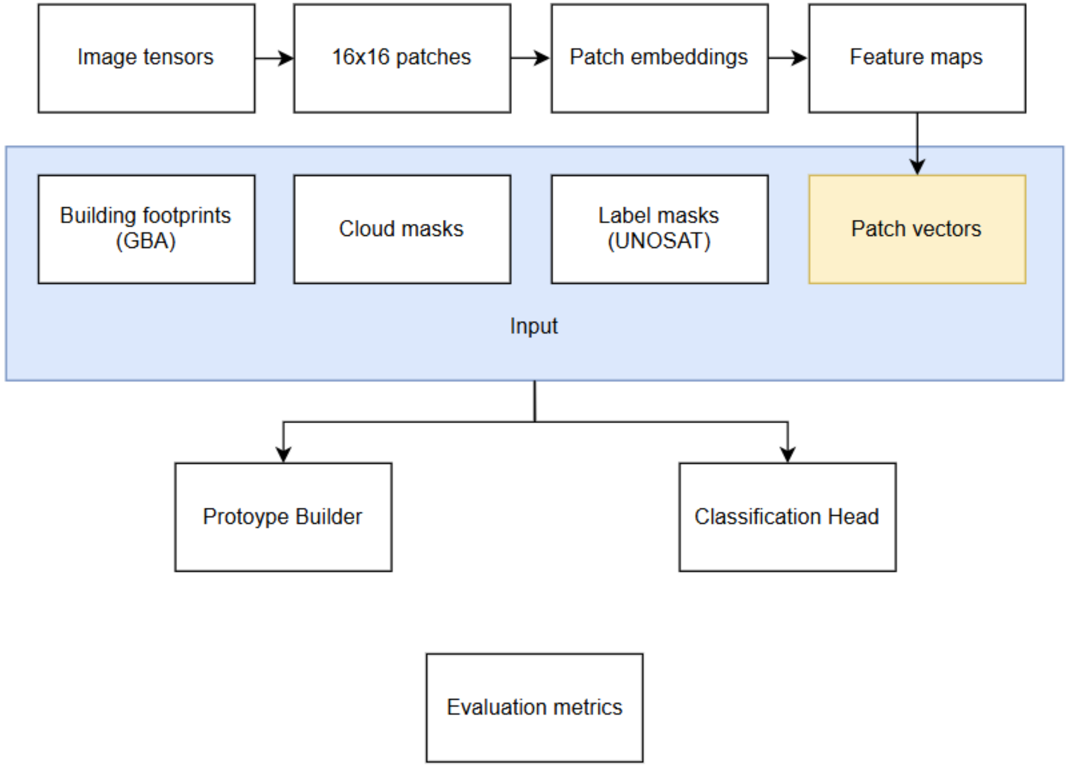
Tiling

Per-Image Quality Checks

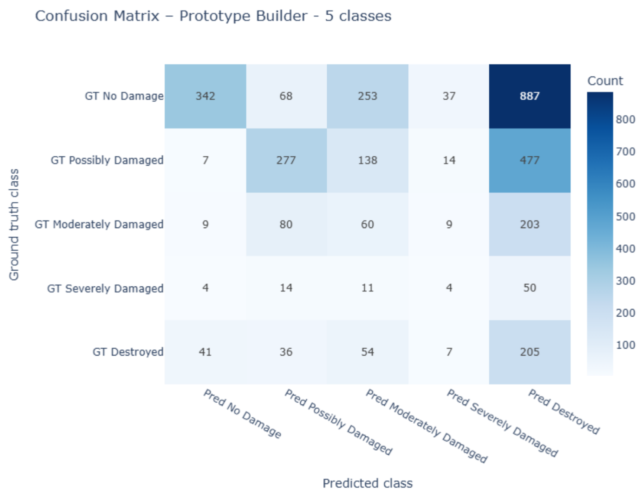
Export

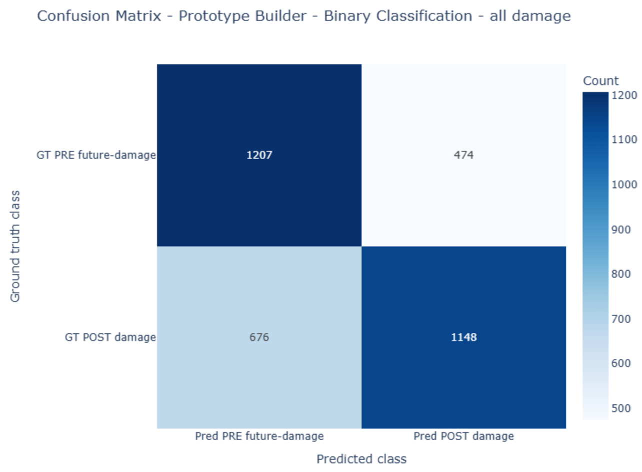
Architecture Pipeline

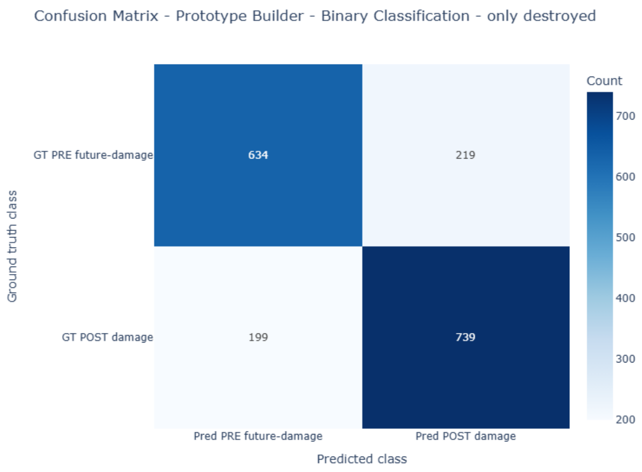
Kompetenzzentrum Krisenfrüherkennung



Results Patch Level – Prototype Builder





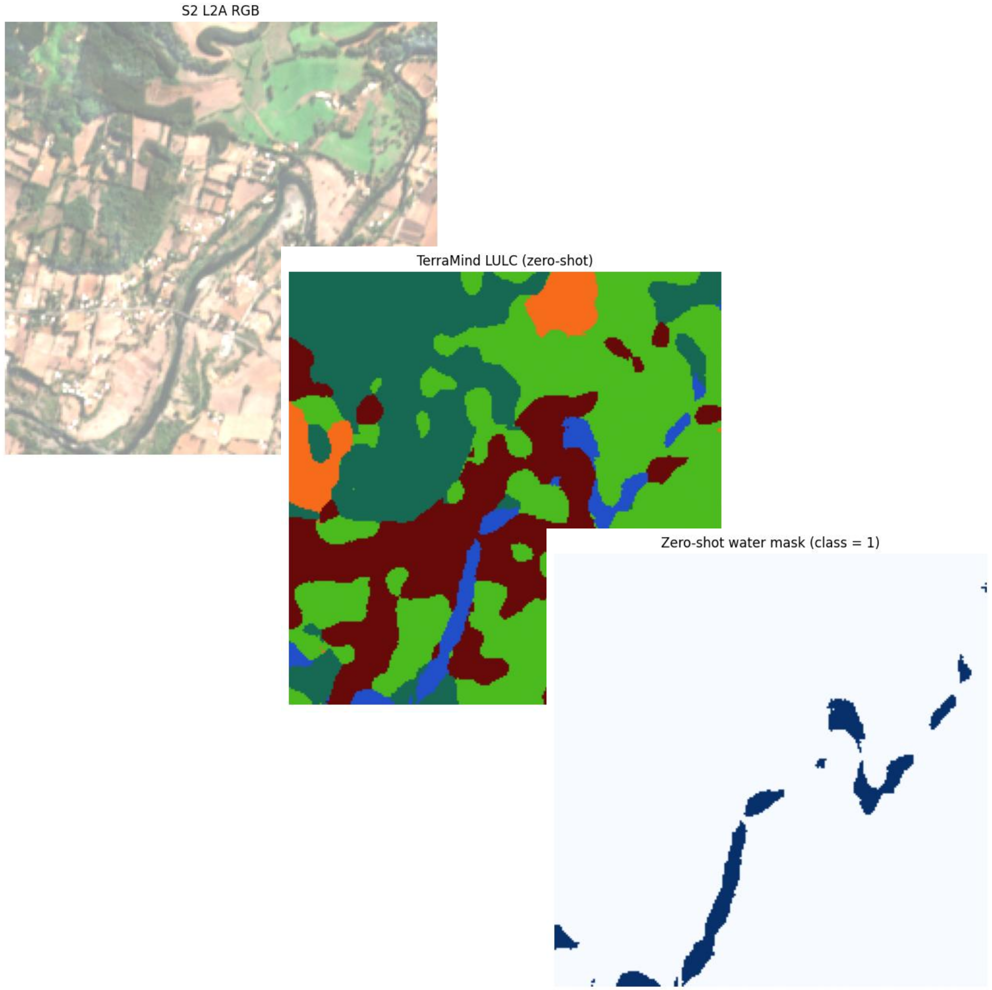


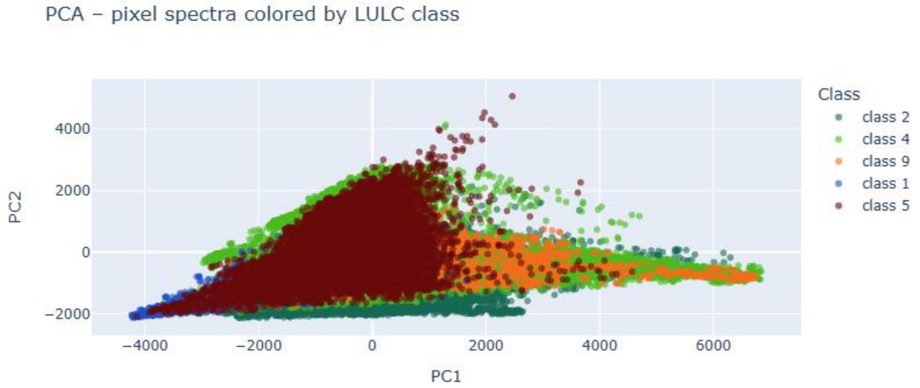
Joint PATCH mIoU(all 0–4) = 0.132 Joint PATCH mIoU(damage1–4) = 0.114

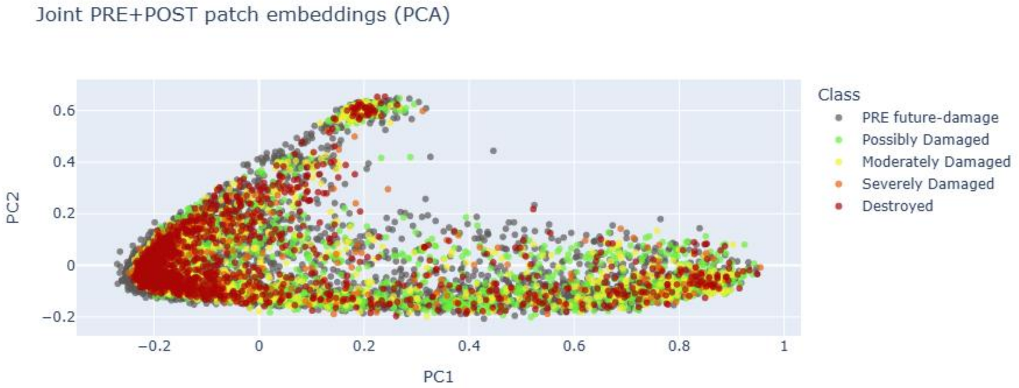
Mean IoU=0.506, Mean F1=0.672

Mean IoU=0.621, Mean F1=0.766

Data Exploration – Patch Level







Limitations – Patch Label Construction

4x4 patch example

	building no label	building no label	building class 1
	building class 4	building class 4	building class 1
	building class 4	building class 4	building class 4

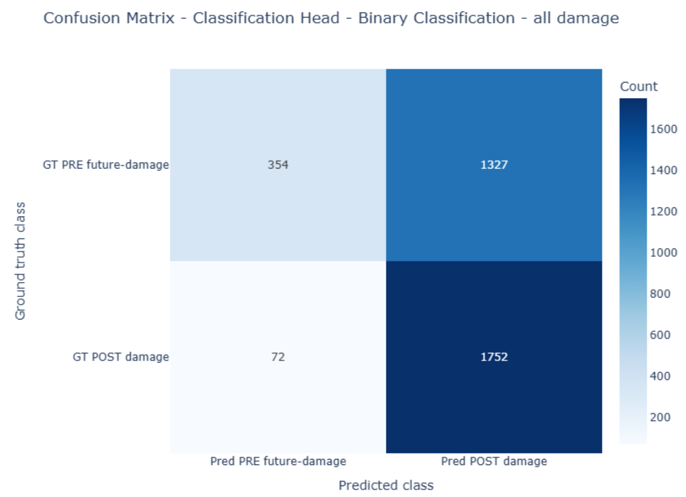
			building class 4

		building class 1	building class 1
		building class 4	building class 4

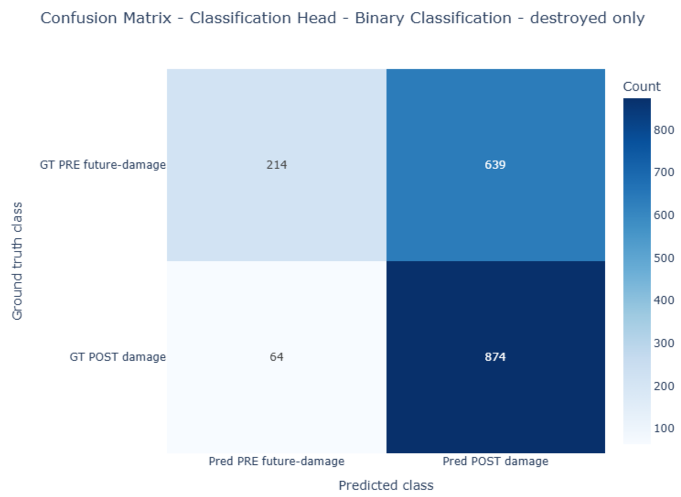
	building class 1	building class 1	building class 1
building class 1	building class 1	building class 1	building class 1
building class 4	building class 4	building class 4	building class 4
building class 4	building class 4	building class 4	building class 4

Results Patch Level – Classification Head

Classification Head -5 classes WIP



Mean IoU=0.348, Mean F1=0.483



Mean IoU=0.371, Mean F1=0.515

Next Steps

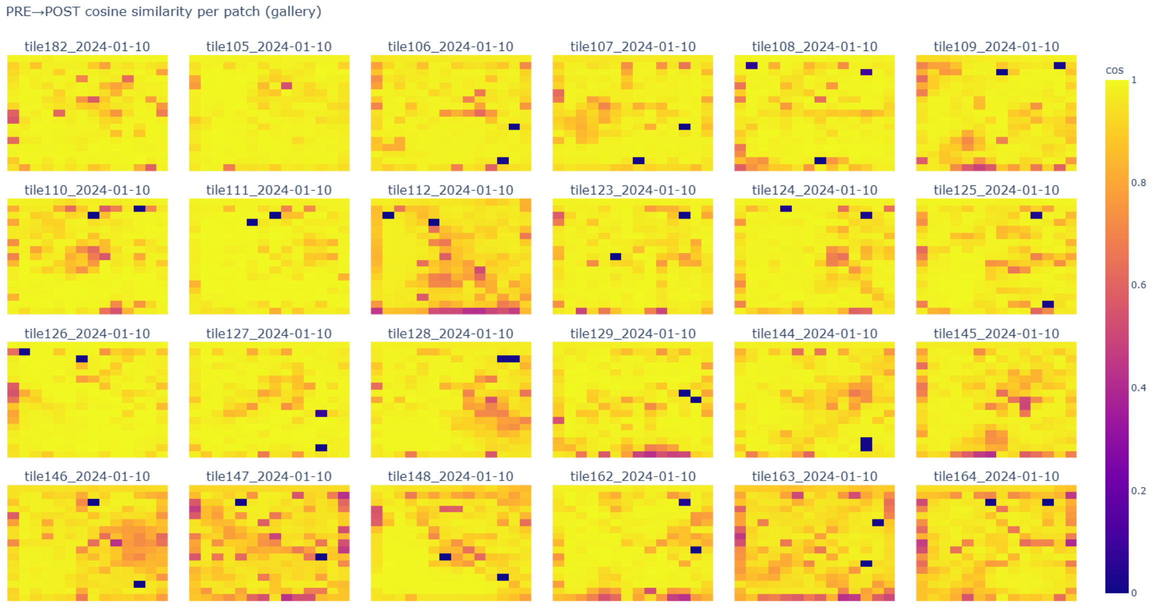
- Improve the classfication head results
- Add the temporal dimension
- Test different image types
- Explore other foundation models

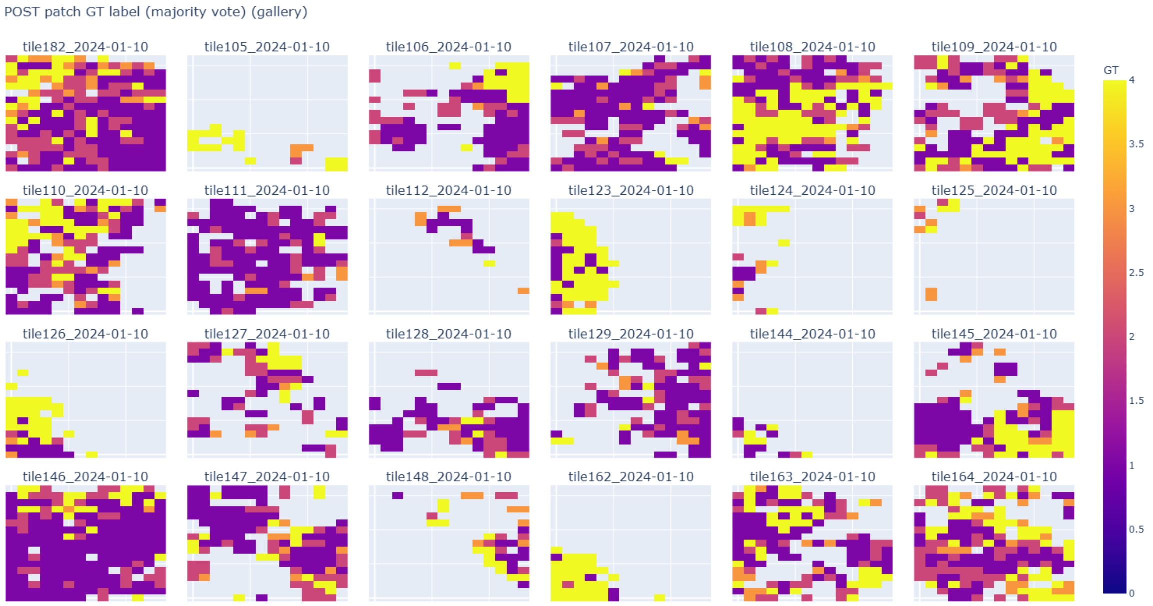
Universität der Bundeswehr München

Kompetenzzentrum Krisenfrüherkennung

Thank you for your attention! ☺

PRE-POST Cos-Sim – Patch Level

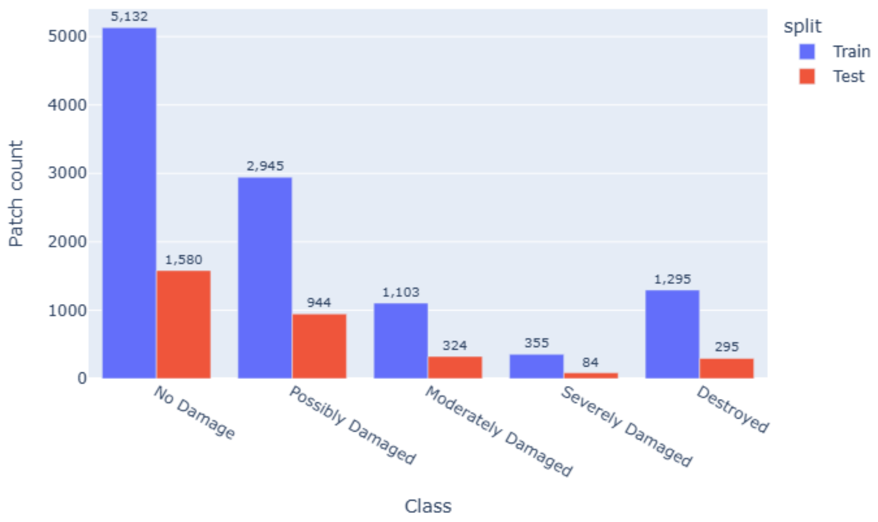




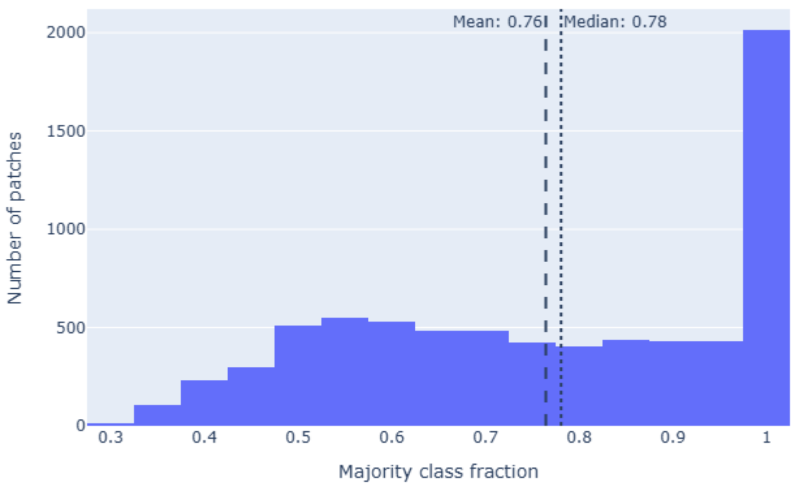
Data exploration

Kompetenzzentrum Krisenfrüherkennung

Patch count accross splits and classes



Patch purity (majority class fraction)



Building pixels

Kompetenzzentrum Krisenfrüherkennung

What happens if we remove patches with few building pixels?

	split	threshold	patches_removed	maj_0	maj_1	maj_2	maj_3	maj_4	lost_px_0	lost_px_1	lost_px_2	lost_px_3	lost_px_4	lost_px_total	patches_removed_frac	lost_px_0_frac	lost_px_1_frac	lost_px_2_frac	lost_px_3_frac	lost_px_4_frac
0	train	10	209	0	14	20	25	150	0	63	93	98	730	984	0.034330	0.0	0.000193	0.000769	0.003262	0.009113
1	train	50	1042	0	152	164	140	586	0	2716	2427	1951	8471	15565	0.171156	0.0	0.008309	0.020062	0.064936	0.105746
2	train	100	1989	0	421	384	242	942	0	13792	10417	6268	25263	55740	0.326708	0.0	0.042194	0.086109	0.208620	0.315366
3	test	10	13	0	1	0	1	11	0	2	0	5	60	67	0.010342	0.0	0.000028	0.000000	0.000614	0.002875
4	test	50	83	0	14	15	12	42	0	207	225	180	496	1108	0.066030	0.0	0.002899	0.007349	0.022108	0.023764
5	test	100	249	0	56	69	27	97	0	1534	1851	982	2726	7093	0.198091	0.0	0.021482	0.060459	0.120609	0.130606
6	train+test	10	222	0	15	20	26	161	0	65	93	103	790	1051	0.030225	0.0	0.000163	0.000613	0.002697	0.007823
7	train+test	50	1125	0	166	179	152	628	0	2923	2652	2131	8967	16673	0.153165	0.0	0.007339	0.017494	0.055804	0.088801
8	train+test	100	2238	0	477	453	269	1039	0	15326	12268	7250	27989	62833	0.304697	0.0	0.038480	0.080928	0.189855	0.277176

Literature Patch Rules

Kompetenzzentrum
Krisenfrüherkennung

- Damage / No damage:
  - Take only completly destroyed buildings (Mueller et al. 2021, , Hou et al., 2024, pixel - Racek et al., 2025)
  - Take only severely and destroyed buildings (Gazel-Anthoine, 2022)
  - Collapse all classes into one (maybe Ballinger, 2024)
- 5 Classes rules:
  - Majority voting (Holail et al. 2024) – after prediction, 50% treshold
  - Max severity (Risso et al. 2024) – at the building level
  - My idea: max severity with an extra percentage treshold – no literature backup