Deep Learning for Building Damage Assessment

From Siamese networks to foundation models, and what still fails in real disasters

Virgile Blua, Tues 3 Nov 2026

A bit of history

Year	Work	Contribution
2017	Fujita et al.	CNNs on aerial pairs, Japan tsunami of 2011
2018	Caye Daudt et al.	Siamese encoder-decoder for change detection
2019	Google + WFP, Xu et al.	CNNs on satellite pairs, tested across earthquakes
2019	xBD, xView2 challenge	850,736 labelled buildings, the shared benchmark
2023	SAM, DINOv2, Prithvi	Pretrained foundation models become available
2024–26	DAVI, HASTE	Damage mapping with few or no target labels

Sources: Fujita 2017; Caye Daudt 2018; Xu 2019; Gupta 2019; DAVI; HASTE

Google, 2019

- Step 1: an object detector finds each building.
- Step 2: a CNN scores it from 0 (undamaged) to 1 (damaged).
- Input: 161 × 161 px crops at 0.3 m, before and after, 6 channels in total.
- Network: four AlexNet-based variants.
- Labels: UNOSAT manual assessments.

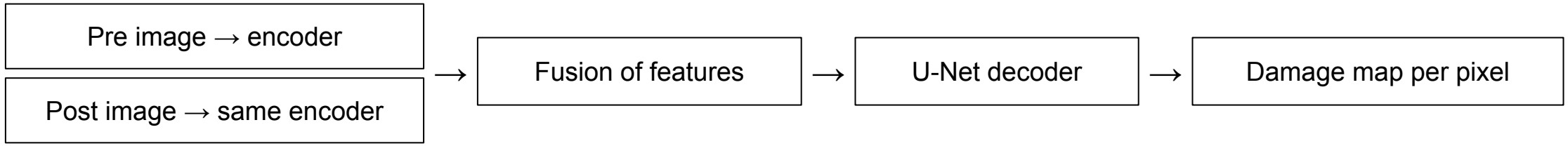
Variant (Haiti 2010)	AUC
Post image only	0.80
Concatenated channels	0.80
Twin tower, concatenate	0.81
Twin tower, subtract	0.83

Key result: processing each date in its own tower and subtracting the features worked best.

Key limit: trained on Haiti and tested on Mexico 2017, AUC falls to 0.62. Adding Indonesia and 10% of Mexico lifts it to 0.76.

Sources: Xu et al., 2019; Google Research blog, 2020

Siamese networks to capture change



- Shared weights: both dates pass through the same filters, so features are comparable.
- Fusion: concatenation, difference or attention between the two feature maps.
- Output: building mask, then no, minor, major damage or destroyed.
- Training: often localization first, then damage. Focal and Dice losses handle class imbalance.
- xView2 winner: four models and 24 inference passes per input. Microsoft later used a single multitask model.
- Limit: later work has rarely beaten the top five solutions on xBD.

Sources: xView2 solution, 2021; Gholami et al.; U-BDD; xView2 code; xFBD 2022

Foundation models: what they add

Large pretrained vision models are reused, so a new disaster no longer needs a full training set.

Model	Type	Use for damage assessment
SAM	Segmentation	Building score on pre and post; a drop flags destruction
DINOv2, DINOv3	Self-supervised features	One embedding per building, then a small classifier
Prithvi	Geospatial pretrained	Encoder for satellite pre and post pairs

- Few labels: macro AUC 0.91 with 10% of labels, vs 0.88 for a ResNet-50 on all labels (xBD).
- No target labels: F1 0.42 on an unseen hurricane with SAM, vs 0.09 for the source model.
- Fast: retraining and re-scoring in seconds, while an analyst adds labels.
- Not automatic: on unseen events at 10 m, Prithvi reached F1 0.63, vs 0.71 for a Siamese U-Net.

Sources: HASTE, Microsoft 2026; DAVI, Ahn 2024; xBD-S12, 2025

The need for a real-use dataset

Why today's benchmarks are not enough

- xView2 test locations also appear in training, so scores flatter the models.
- ChangeMamba falls from F1 0.80 to 0.66 on unseen events (xBD-S12).
- An xBD-trained model reaches binary F1 0.25 on the Ahr flood, 2021, partly due to noisy labels.
- In the xBD flood subset, 91.7% is no damage and 0.33% is destroyed.

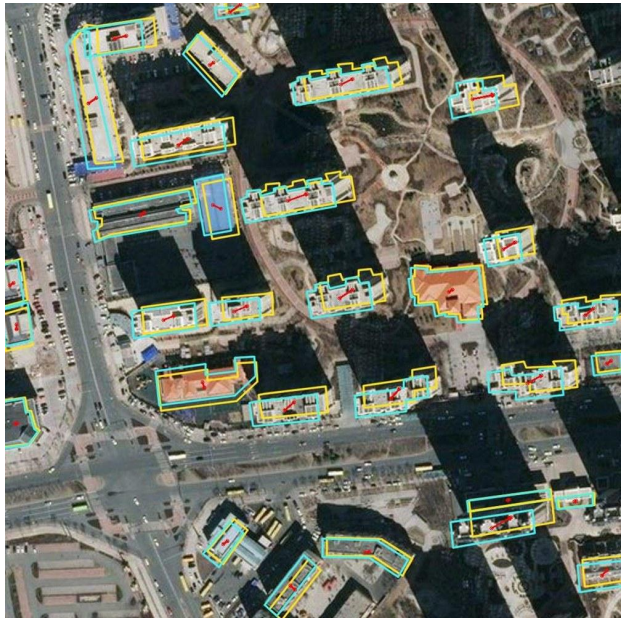
What it must contain: pre- and post-image pairs with

- Different satellites and viewing angles, and mismatched resolutions
- Flooded buildings with intact roofs, verified from the ground
- Buildings swept away, labelled destroyed from the pre-image
- Months between dates, with seasons and unrelated construction

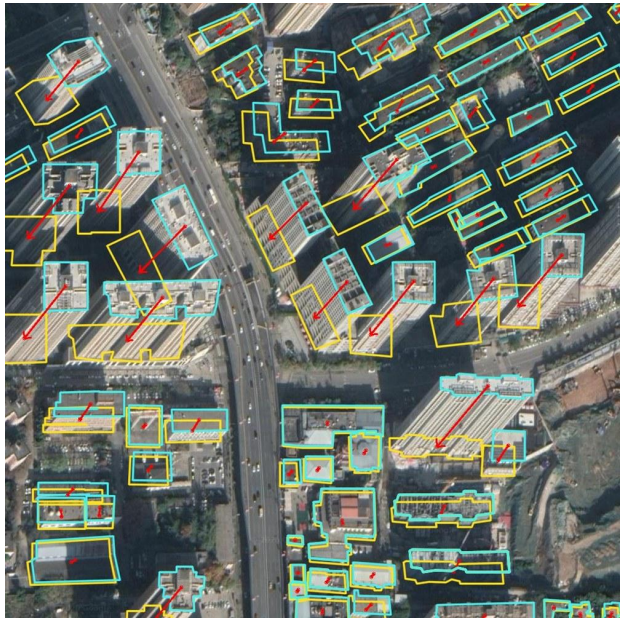
Test split: by event and region, so the disaster is new to the model.

Sources: Gerard 2024; xBD-S12 2025; AB2CD 2023; SPAUNet 2024

Challenge 1: Perspective



Harbin: small offset



Xi’an: large offset, façades visible

- Seen at an angle, a high roof shifts away from its footprint.
- Offset length depends mainly on height, its direction on the view.
- For the model: pre and post pixels no longer align, so it sees false change.
- Angle gaps reach 22° in Spain 2024 and 29° in Nepal 2024.
- Leads: shift augmentation, view conditioning, roof + offset prediction.

Image origin: BONAI figure (TPAMI 2022), cropped by us; mainly Google Earth and Microsoft Virtual Earth, 0.3–0.6 m. Blue: roof, yellow: footprint, red: offset. Two different places, not a pre/post pair.

Sources: BONAI 2022; ORNL 2024; onboard BDA 2026; Maxar catalog metadata

Perspective: a real pair, Porto Alegre 2024



Pre, 12 Dec 2022, 11.0° off-nadir



Post, 9 May 2024, 42.6° off-nadir

- Same place, 31.6° of angle gap: towers lean and façades appear.
- In the post-image, towers hide ground that is visible in the pre-image.
- Other changes: 0.32 m vs 0.74 m, two satellites, 17 months apart.
- For the model: the towers look changed because of the view, not the flood.

Image origin: Maxar Open Data Program, event Brazil-Flooding-May24, tile 213131133011, WorldView-3 (pre), GeoEye-1 (post), CC BY-NC 4.0. Cropped by us, 520 m wide, Porto Alegre, Brazil, 30.047° S, 51.231° W.

Sources: Maxar Open Data Program; Maxar catalog metadata

Challenge 2: Resolution



0.42 m, original



0.75 m, simulated by us



3 m, simulated by us

- Real case: 0.32 m pre and 0.75 m post on one Nepal 2024 tile, a factor of 2.3.
- For the model: the same roof covers fewer pixels, so features differ even with no damage.
- Binary damage needs a post-image finer than about 3 m; four damage levels need 1 m or finer.

Image origin: Maxar Open Data Program, event Libya-Floods-Sept-2023, GeoEye-1, 13 Sep 2023, tile 120200213130, CC BY-NC 4.0. Cropped near Derna, Libya, 32.767° N, 22.635° E.

Sources: AB2CD, Nitsche 2023; Maxar catalog metadata

Challenge 3: Flooded, but the roof looks fine



Pre, 1 Jul 2023



Post, 13 Sep 2023

- From above, roofs look intact, but water enters at ground level.
- For the model: outlines stay the same and few pixels change.
- Seen in Valencia 2024, Typhoon Yagi 2024 and Cyclone Senyar 2025.
- Leads: use flood extent as context, and ground-verified labels.

Image origin: Maxar Open Data Program, event Libya-Floods-Sept-2023, GeoEye-1, tile 120200213130, CC BY-NC 4.0. Cropped near Derna, Libya, 32.767° N, 22.635° E. Shows mud after the water receded.

Sources: SPAUNet 2024; xFBD 2022; EMSR773; Yagi; Senyar

Challenge 4: Swept away, nothing left



Pre, 1 Jul 2023



Post, 13 Sep 2023

- In Nepal, Melamchi 2021 and Rasuwa 2026 washed away buildings and roads.
- For the model: no roof and no rubble in the post-image, so no visual cue.
- Expected logic: in the pre-image and gone in the post-image means destroyed.
- Leads: pre-event footprints, and disappearance detection with SAM.

Image origin: Maxar Open Data Program, event Libya-Floods-Sept-2023, GeoEye-1, tile 120200213130, CC BY-NC 4.0. Cropped near Derna, Libya, 32.764° N, 22.644° E. Not an image of Nepal.

Sources: Al Jazeera 2026; Sentinel Asia 2021; Ahmad and Saha 2024; DAVI 2024

Challenge 5: Time between images



Pre, 22 Nov 2022, 25.1° off-nadir



Post, 6 Oct 2024, 42.1° off-nadir

- Here, 684 days apart: dry fields turn green and most long sheds are gone.
- Median gap on this event's tiles: 267 days.
- Building to trees: razed and replanted, unrelated to the disaster.
- Field to building: new, damaged or flooded, with nothing to compare.
- For the model: separate disaster change from normal change.

Image origin: Maxar Open Data Program, event Nepal-Floods-Sept-2024, tile 031331122023, WorldView-3 (pre), GeoEye-1 (post), CC BY-NC 4.0. Cropped by us, 480 m wide, near Bhaktapur, Nepal, 27.672° N, 85.464° E.

Sources: Maxar Open Data Program; gaps computed by us from catalog metadata

Limit of satellites: cracks in walls



Post, 13 Feb 2023, 170 m wide



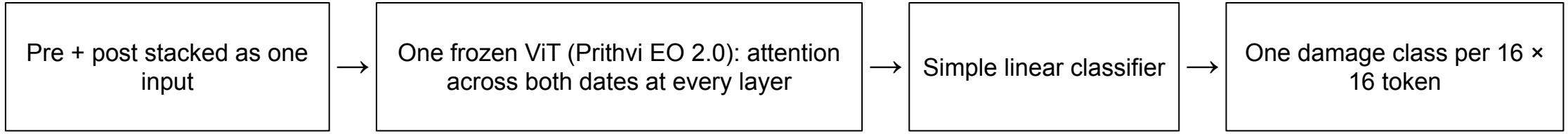
Zoom, 45 m wide: no crack visible

- Satellite, 41° at 0.53 m: debris shows, cracks do not. A 1 cm crack is 50× smaller than a pixel.
- Oblique aerial, 0.14 m: small cracks remain challenging.
- UAV: close views can show cracks, but cover small areas.
- InSAR: tracks surface and height change, not single cracks; useful above EMS-98 level 2.

Image origin: Maxar Open Data Program, event Kahramanmaras-turkey-earthquake-23, tile 031131323022, WorldView-3, 41° off-nadir, CC BY-NC 4.0. Cropped by us, Gölbaşı, Turkey, 37.787° N, 37.646° E.

Sources: Oblique façade study; Crack detection at several scales; Kumamoto InSAR; InSAR thesis

SSTA: a single-stream network



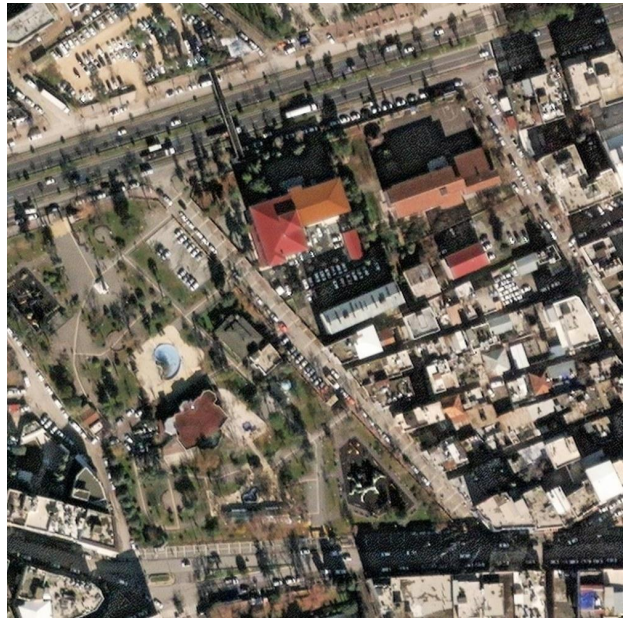
	Siamese models	**SSTA**
Input	Two parallel branches	One joint input
Fusion	Late, with a change decoder	Native, at every layer
Output	Pixel mask or object detection	Token classes, 8 m cells at 0.5 m
Trained parameters	32.5 M to 87.8 M	1.5 M (DoRA), 4.7 M (LoRA)

xBD F1, no / minor / major / destroyed: SSTA LoRA 0.938 / 0.581 / 0.765 / 0.849; MambaBDA 0.959 / 0.627 / 0.765 / 0.885. Both SSTA variants beat ChangeOS-101 and DamFormer on every class.

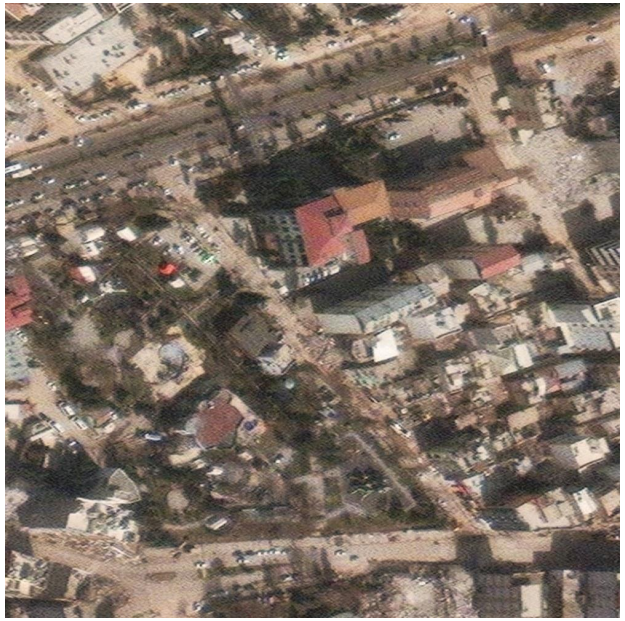
The model is a proof of concept, trained only on the xBD dataset to see the potential of such networks.

Source: Blua, Cioppa, Van Droogenbroeck and Ceccato, “Single-Stream Temporal Attention for Building Damage Assessment”, 2026

Why tokens suit this challenge



Pre, 18 Jan 2023, 6.4°, 0.47 m



Post, 28 Feb 2023, 47.5°, 0.66 m

- Same place, 41° angle gap: the red roof shifts roughly 20 m and façades appear.
- Pixel model: must trace exact masks through blur, debris and hidden ground. This looks unrealistic.
- Token model: only asks whether this 8 m cell shows damage. No outline to draw.
- Swept away (challenge 4): no building is left, so what should a pixel model mask? A token only flags the cell.
- Responders act at area or block level, so cells stay useful.

Limits: a shift of 2–3 cells still matters, there is no building identity, and these scenes are untested.

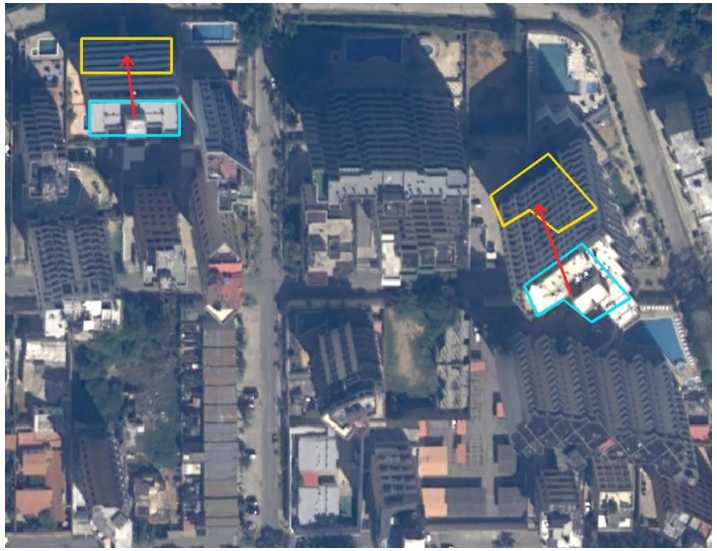
Image origin: Maxar Open Data Program, event Kahramanmaras-turkey-earthquake-23, tile 031131332211, WorldView-2 (pre), WorldView-3 (post), CC BY-NC 4.0. Cropped by us, 300 m wide, Adıyaman, Turkey, 37.761° N, 38.269° E.

Source: Blua et al., 2026. The comparison on this scene is our reasoning, not a measured result.

First test: Venezuela earthquakes, June 2026



Post, wide view; white box: close-ups



Pre, 5 Jan 2026: towers lean strongly



Post, 26 Jun 2026: same outlines, cells

- Trained on xBD alone, not trained yet directly to solve these challenges.
- Hard input: 5.5 months apart, two satellites, strong lean in the pre-image.
- Good: cells gather around debris, with few false alarms on standing towers. In a disaster, better no prediction than a wrong one.
- To fix: here cells partly follow the pre-image roof (blue), not the debris at the footprint (yellow); some debris is missed.
- A good way forward: retrain on real pairs that contain these challenges, then score against ground truth.

Image origin: our viewer, Copernicus EMS EMSR884 AOI12, La Guaira coast near Caracas, 10.61° N, 67.01° W. Pre: Pléiades Neo. Post: Legion. Outlines drawn by hand by us, approximate: blue roof, yellow footprint, red offset. Source: Blua et al., 2026; Copernicus EMS.

Data sources to build it

Source	Content	Access	License
Maxar	Optical, 0.3–0.8 m	s3://maxar-opendata/	CC BY-NC 4.0
Capella	Radar (SAR)	s3://capella-open-data/data/	CC BY 4.0
Umbra	Radar (SAR)	s3://umbra-open-data-catalog/	CC BY 4.0
NOAA ERI	Aerial, US events	s3://noaa-eri-pds/	Open, credit asked
Sentinel-1/2	Radar, optical, 10 m	Copernicus Data Space	Free and open
xBD	850,736 labelled buildings	xview2.org, registration	See site
BRIGHT	14 events, optical and SAR	Zenodo	Varies by event
Noto 2024	140,208 footprints, labels	Zenodo	CC BY 4.0
Copernicus EMS	Expert damage grading maps	mapping.emergency.copernicus.eu	Free, cite; some restricted
Google tiles	Satellite tile service	Map Tiles API	No storage or derivatives
Mapbox tiles	Satellite tile service	mapbox.satellite, Raster Tiles API	See terms
Esri tiles	World Imagery, dated in Wayback	ArcGIS Living Atlas	See terms

Coupling: open imagery + Copernicus grading + pre-event footprints (OpenStreetMap, Microsoft, Google). Tile services help as pre-event views, but capture dates vary and terms can forbid reuse for training.

Sources: Maxar; Capella; Umbra; NOAA; BRIGHT; Noto; Copernicus terms; Google terms, via Cesium; Mapbox; Esri Wayback