LIÈGE université

VIULab
Vision and Information Understanding Laboratory

JRC Workshop - AI for Building Damage Assessment

Day 2 - 11 November 2026 - Ispra

The Three Pillars of the Building Damage Assessment Task

Task definition, data, and data quality

Anthony Cioppa - University of Liège

Joint work with V. Blua, S. Piérard, P. Ceccato, M. Van Droogenbroeck



www.viulab.be

JRC EUROPEAN COMMISSION

Emergency Management

The authors of today’s presentations



Anthony Cioppa

Professor, ULiège (VIULab)

Now



Virgile Blua

Independent researcher

10:30
Day 3 hands-on



Sébastien Piérard

Postdoc researcher, ULiège



Pietro Ceccato

JRC, Copernicus EMS

Moderator



Marc Van Droogenbroeck

Professor, ULiège (VIULab)

14:30

A collaboration between **computer vision** research and **operational emergency mapping.**

Today's topic: revisiting the task of BDA

Overview

Revisiting the Task of Building Damage Assessment by Satellite Imaging

A. Cioppa, V. Blua, S. Piérard, P. Ceccato, M. Van Droogenbroeck

Book chapter - arXiv:2610.13857 (2026)

82 pages - 160 references - reading material of the workshop

Scan for the arXiv preprint version

arxiv.org/abs/2610.13857



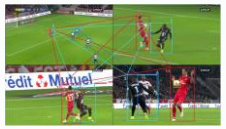
What defining tasks in computer vision taught us

1 Defining tasks


Action spotting


Camera calibration


Player tracking


Re-identification

More than 20 tasks defined in SoccerNet, each with its own **data**, **annotations**, and **evaluation**.

2 Annotating at scale


scientific data

OPEN

DATA DESCRIPTOR

Scaling up SoccerNet with multi-view spatial localization and re-identification

SoccerNet: 20M+ manual annotations on 850 hours of videos, published in journals and conferences.

Nature Scientific Data, CVPR, CVsports, Sports Engineering, MMSports,…

3 Open challenges and community



- **3,000+ submissions** from 750 teams over 6 years
- **2,000+** members on Discord
- Co-organisers of the **CVsports** workshop at CVPR

4 Evaluating fairly

Performance-based ranking (Piérard et al., CVPR 2025) and **the Tile** (Piérard et al., 2025): how to compare methods soundly.

More on this → **M. Van Droogenbroeck, 14:30**

Our lessons on defining the task, the data, and the evaluation.

SoccerNet is a joint effort with KAUST (S. Giancola, B. Ghanem) and many contributors.

Table of content

JRC Workshop · Day 2 · 11 November	
09:00	**The three pillars of the BDA task** A. Cioppa
09:40	BDA: data quality A. Cioppa
10:30	Deep learning for BDA V. Blua
14:30	Performance-based ranking of BDA methods M. Van Droogenbroeck
Day 3	Hands-on: algorithms on VHR images V. Blua



Why revisit BDA?

- A decade of research, no shared foundation
- The three pillars
- The BDA task card

What should we predict?

- One task, many names
- Four candidate formulations
- Our proposal: building instances

From which data?

- Modality, time and geography
- Generalization and data splits
- Existing datasets

Venezuela, 24 June 2026: how good were the damage maps?

Motivation

Two earthquakes, M7.2 and M7.5, struck Venezuela.

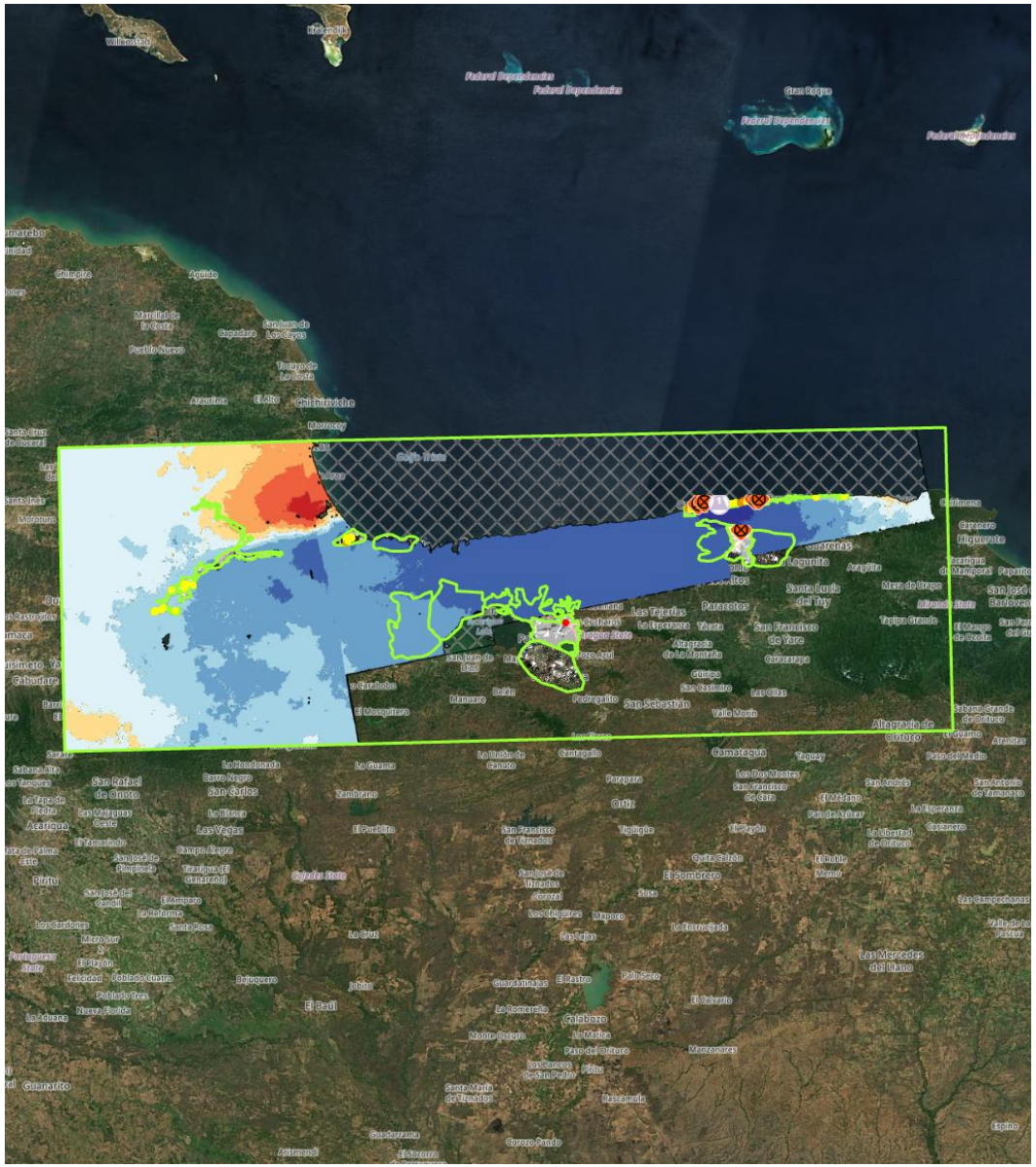
Copernicus EMS was activated (EMSR884): damage mapping over **13 areas of interest**.

Mapped: **3119 potentially affected structures**.

Tests of **AI-driven products** were also carried out[1].

How accurate were they? How can we assess and rank them?

Before answering, we must agree on **what is predicted**, **from which data**, and **against which reference**.



Source: Copernicus EMS, EMSR884 (JRC, 2026).

[1]*Source: Damage analysis of Copernicus Sentinel-1 data by Corey Scher and Jamon Van Den Hoek of Oregon State University, from gis.earthdata.nasa.org*

What you will take away from this talk

Objective

Explain why building damage assessment (BDA) still lacks a shared definition.

Compare the candidate task formulations and what each of them implies.

Describe BDA data along three axes: modality, time and geography.

Identify the main quality issues of the inputs and of the reference annotations.

From expert damage maps to learned models

General objective

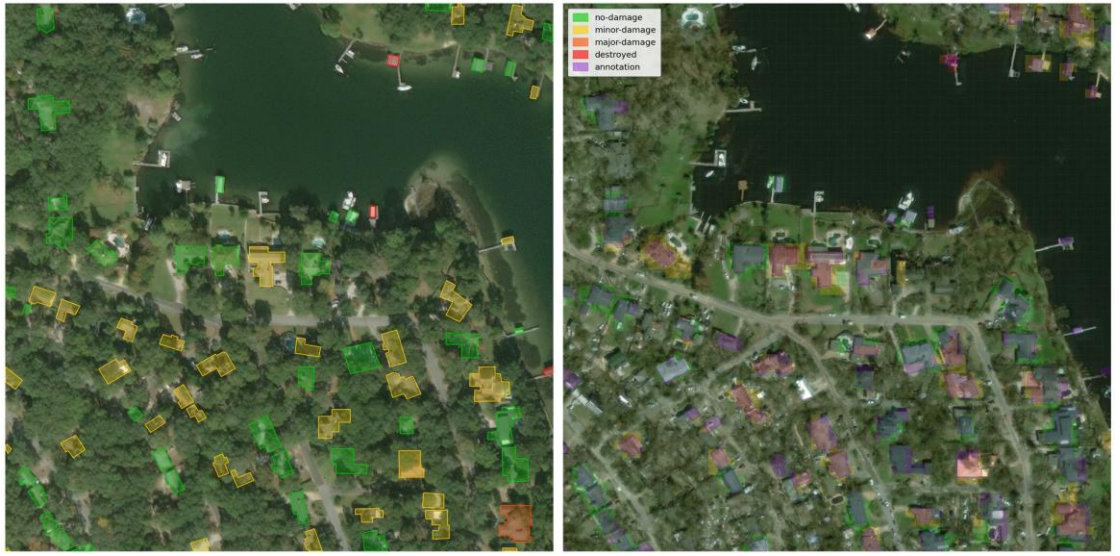
Building damage assessment (BDA): automatically assessing the damage state of buildings from Earth observation data after a disaster.

Overview

Rapid maps are produced by **experts**, through **visual inspection**, which is “slow”, costly, and hard to scale.

Now, **supervised deep learning** has become widely adopted in many fields.

Learning makes data and annotations mandatory.



AI model learned BDA on xBD (Hurricane Michael, 2018)

Three pillar approach of a task

Problem statement

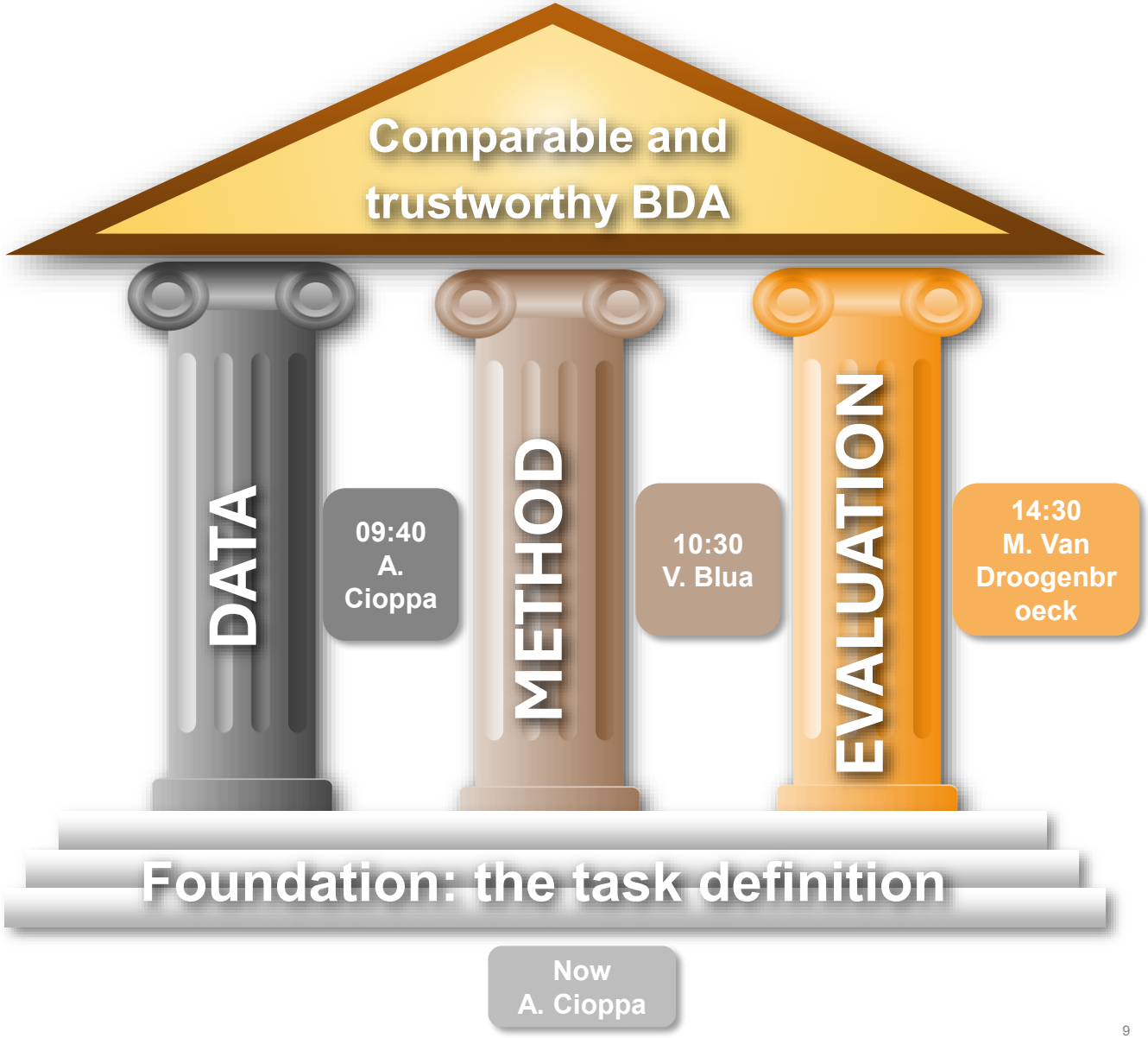
Task: defined in a single non-ambiguous way.

Available data: range of data and annotations that will be used to solve the task.

Method: ranging from manual to AI-based.

Evaluation: methodology that reflects the task objectives.

We revisit BDA from the foundation, on these three pillars.



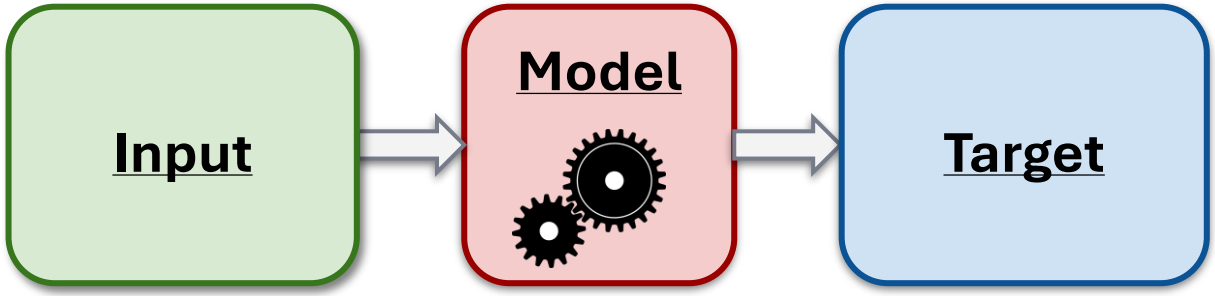
A well-defined task is a clear statement

Principle

A well-defined task must specify its **inputs** and its **prediction target**, so that dataset construction, model design and evaluation can be aligned.

A **scenario** also fixes the constraints on the data and the evaluation protocol.

Today, we fill in this task card together.



BDA task card

Input data	?
Prediction target	?
Reference labels	?
Evaluation	?

Table of content

JRC Workshop · Day 2 · 11 November	
09:00	**The three pillars of the BDA task** A. Cioppa
09:40	BDA: data quality A. Cioppa
10:30	Deep learning for BDA V. Blua
14:30	Performance-based ranking of BDA methods M. Van Droogenbroeck
Day 3	Hands-on: algorithms on VHR images V. Blua



Why revisit BDA?

- A decade of research, no shared foundation
- The three pillars
- The BDA task card

What should we predict?

- One task, many names
- Four candidate formulations
- Our proposal: building instances

From which data?

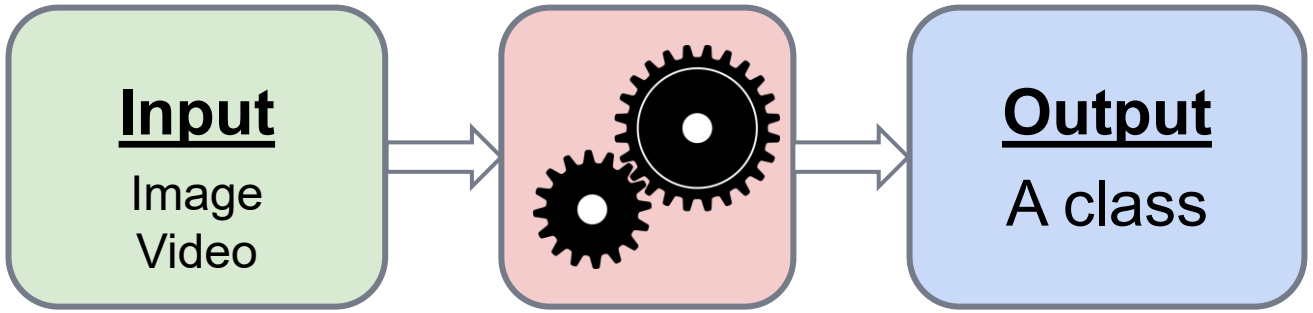
- Modality, time and geography
- Generalization and data splits
- Existing datasets

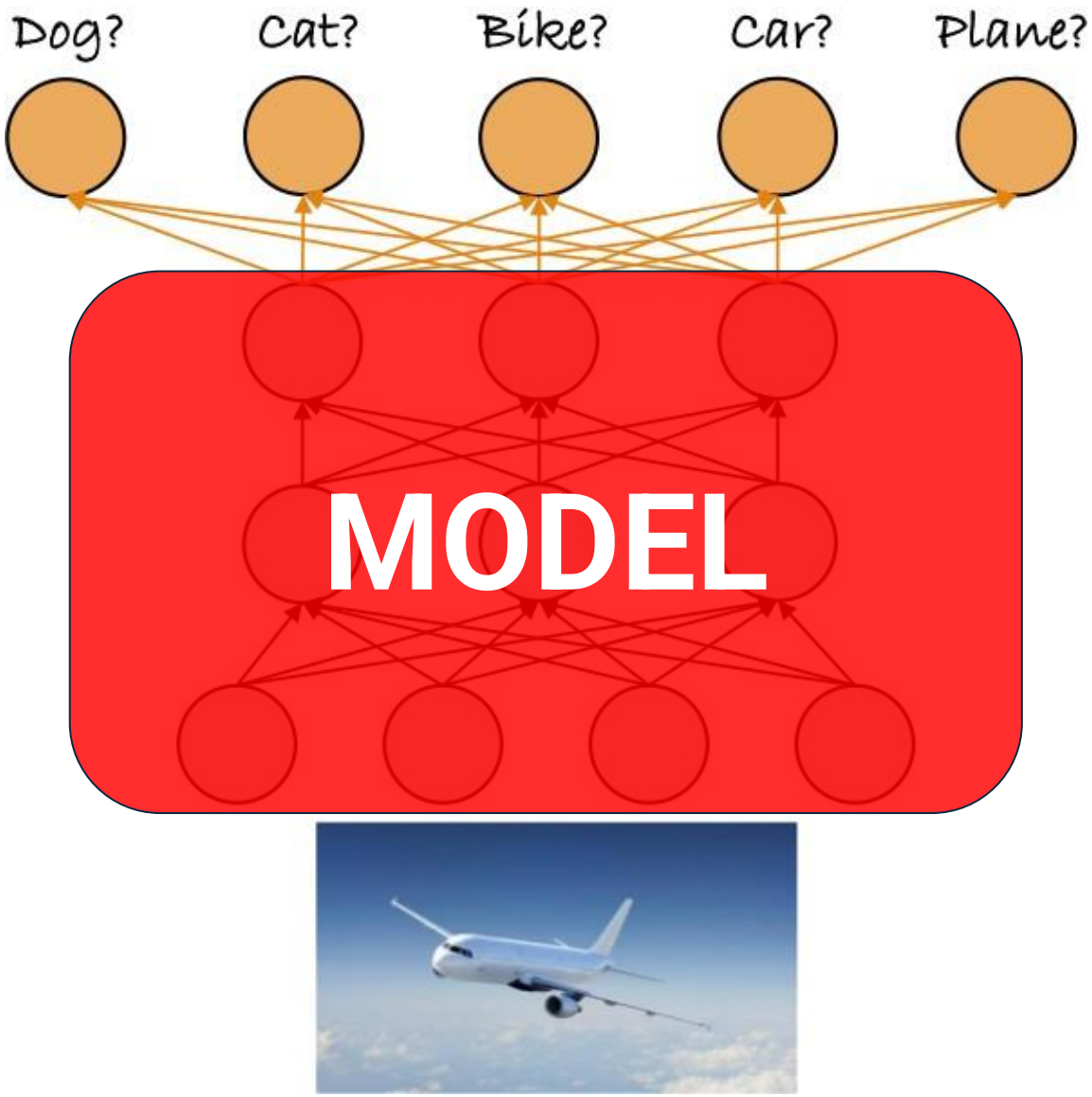
Example of task definitions: classification

Problem statement **Example**

Classification task:

Given an image/video, classify it among one of N classes.



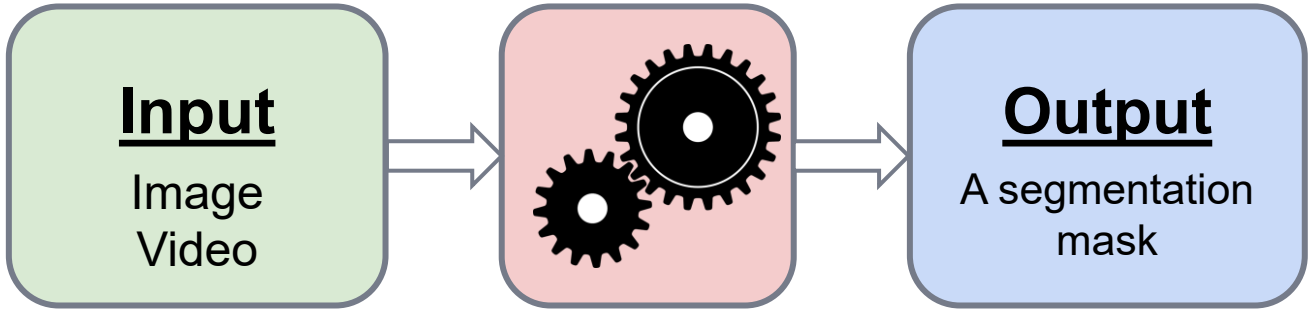


Example of task definitions: segmentation

Problem statement **Example**

Segmentation task:

Given an image/video, classify each pixel among one of N classes.





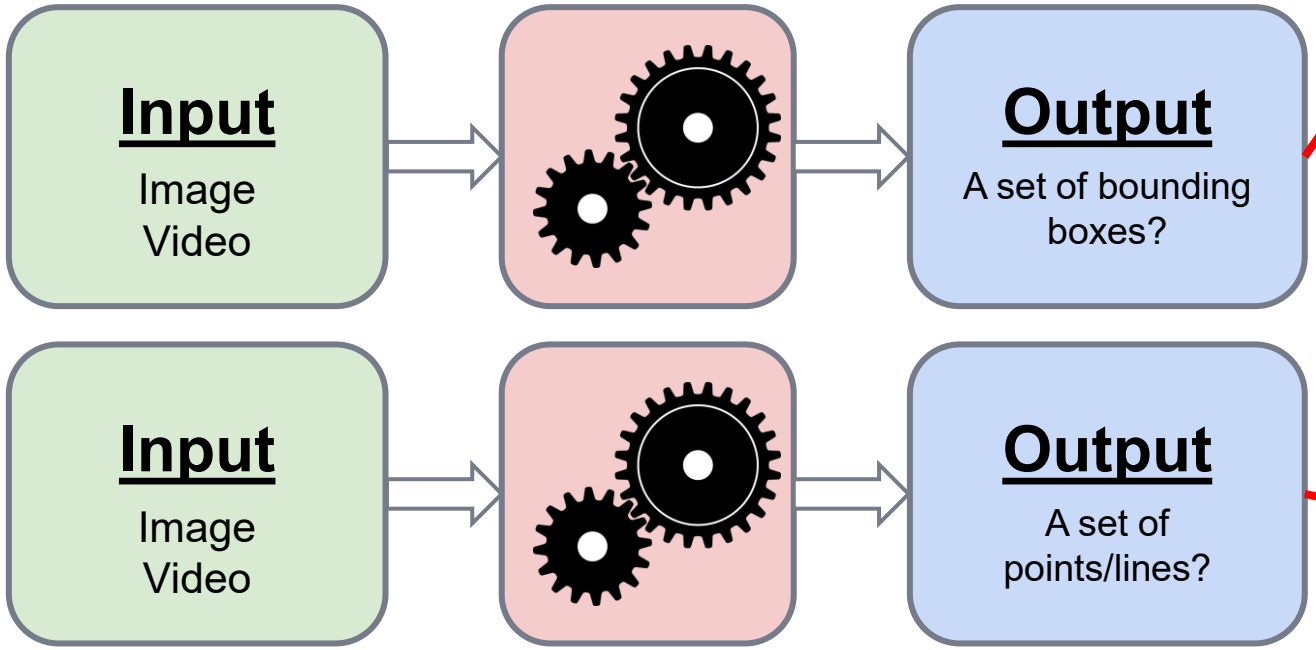
Source: https://towardsdatascience.com/semantic-segmentation-popular-architectures-dff0a75f39d0

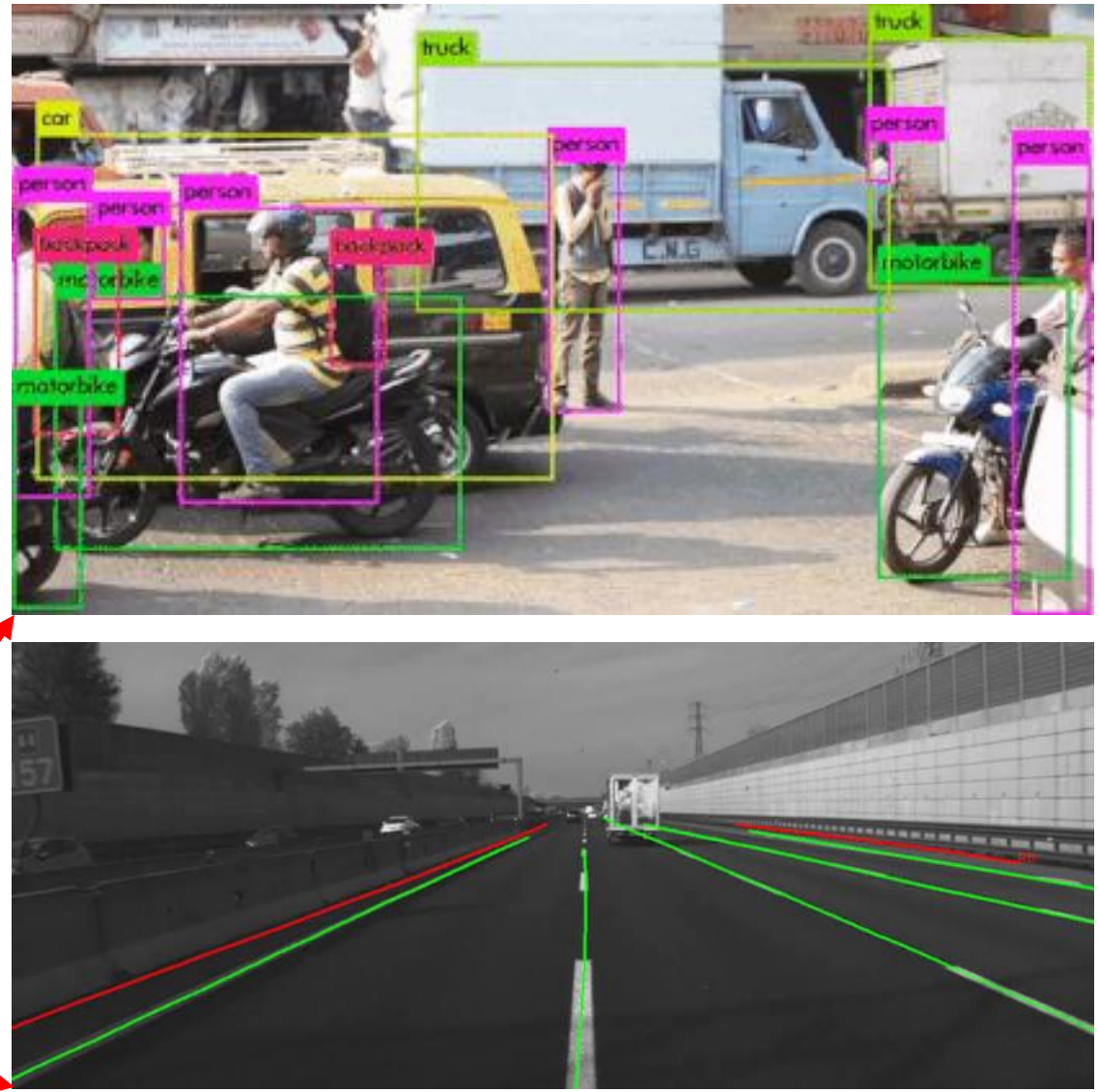
Example of task definitions: detection

Problem statement **Example**

Detection task:

Given an image/video, localize each object and provide a class label.





Sources: https://towardsdatascience.com/review-yolov3-you-only-look-once-object-detection-eab75d7a1ba6 and An Ego-Lane Estimation Method With Sensor Failure Modeling

BDA: One task, many definitions

Examples from the literature

*"This critical task of **identifying** damaged structures is essential for ensuring public safety […]"*

Identifying = detecting? segmenting? classifying? grading?

*"We formulate the **damage detection** task as an **anomaly detection** problem […]. While this task falls within the broader scope of **BDA**, our focus is specifically on **binary classification** […]"*

Four different problems:

Detection
localize every object of interest

Anomaly detection
find outliers of a stable 'normal' distribution

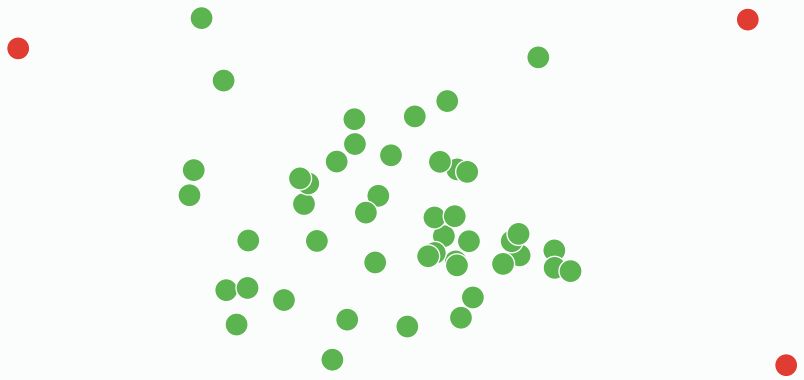
Binary classification
assign one of two classes to each sample

BDA
the umbrella term for the application

A pattern across the field: these terms imply different outputs, annotations and evaluations.

Is a damaged building an anomaly?

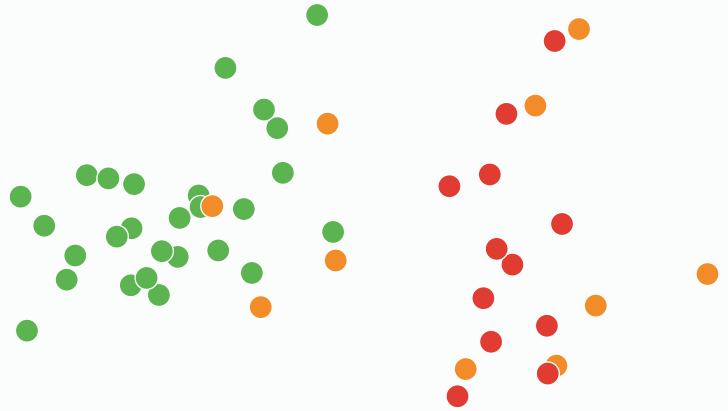
What anomaly detection assumes



A stable ‘normal’ class with rare outliers.

The models learn from normal samples only.

What a disaster produces



A distribution shift: damage is frequent and structured, *e.g.* ~40% damaged buildings

intact damaged destroyed

Class imbalance alone does not make damage an anomaly.

Sun, Wang & Eineder, IEEE GRSL 2024; Chandola et al., ACM CSUR 2009; Fujita et al., MVA 2017.

When the stated task and the implemented method differ

Examples from the literature

Stated task: classify each **building** into four damage levels (JRC guidelines): Destroyed - Damaged - Possibly damaged - No damage

Implemented pipeline:



Limitations

Annotations, model and scores do not refer to the **stated task**, so a comparison with building-level methods is not meaningful.

Without an explicit task statement, results cannot be compared.

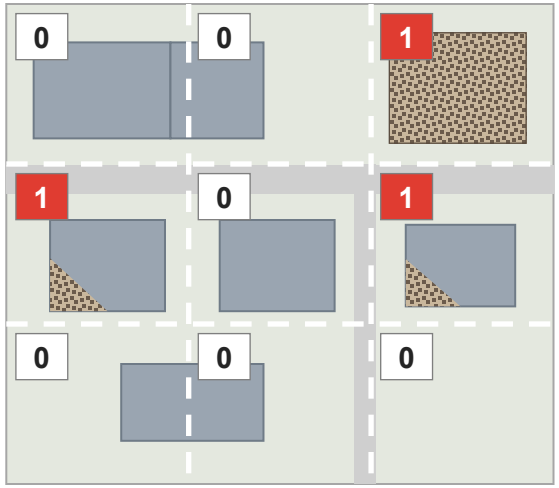
Adriano et al., Remote Sensing 2019, as an example of a pattern found across the literature

Four ways to formulate BDA

Same scene, four different targets.

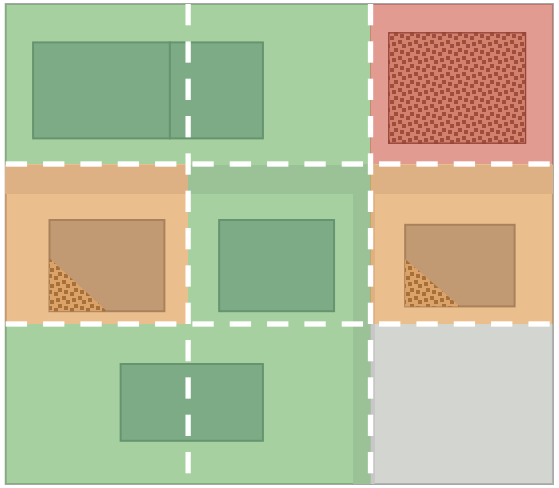
The task formulation decides what a prediction is.

(a) Binary classification



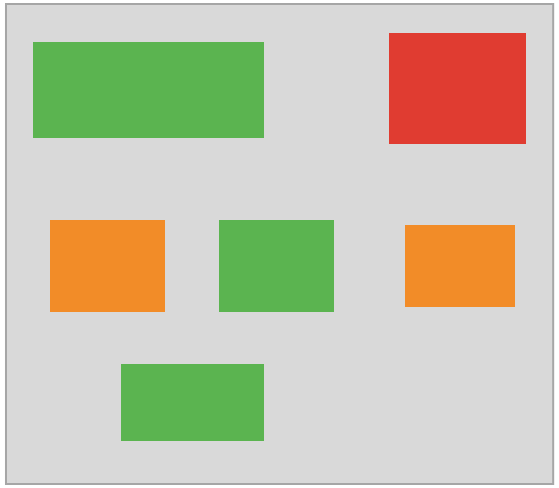
Per patch: is there damage? 1 / 0

(b) Multi-class classification



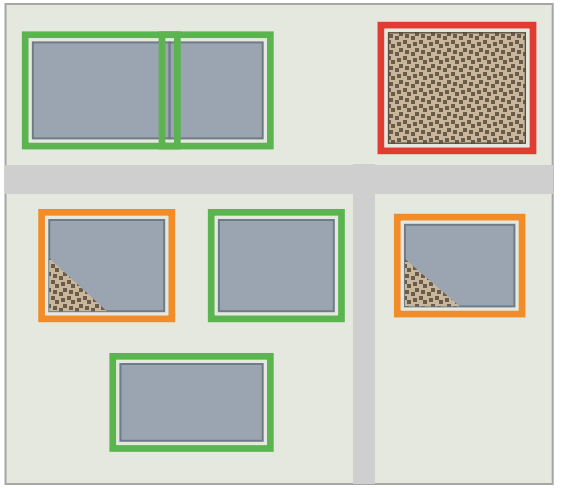
Per crop: one damage class

(c) Semantic segmentation



Per pixel: one class

(d) Detection / instances



Per building: a shape + a damage class

Destroyed | Damaged | No visible damage | No building / background

Classifying image crops: simple, but ambiguous

Properties

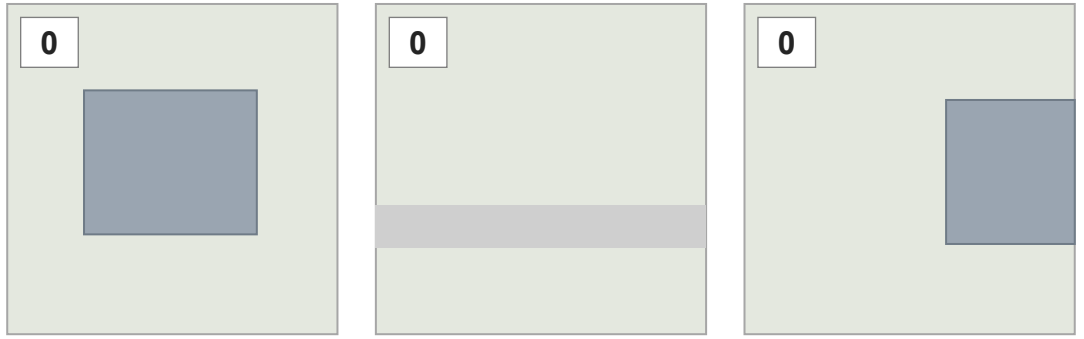
Classification: **Simple**, **cheap to annotate**, and widely used in the literature.

Limitations

The **negative class** mixes different situations.

For densely packed buildings the localization stays **implicit and depends on the patch size**.

Three different situations, one label: 0



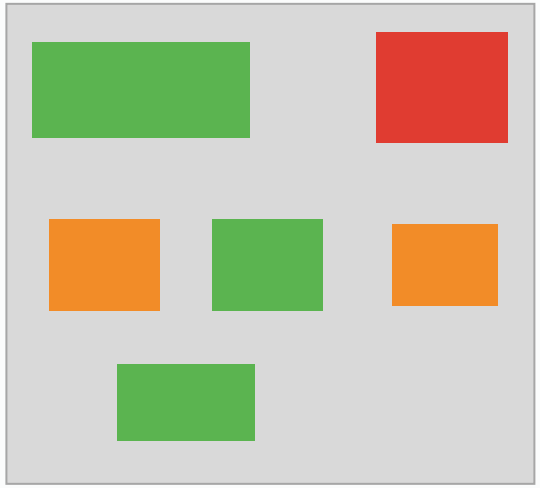
Intact building | No building | Partly visible building

Each patch answers ‘is there damage here?’, not ‘which building is damaged?’

Pixels or instances?

Properties

Semantic segmentation



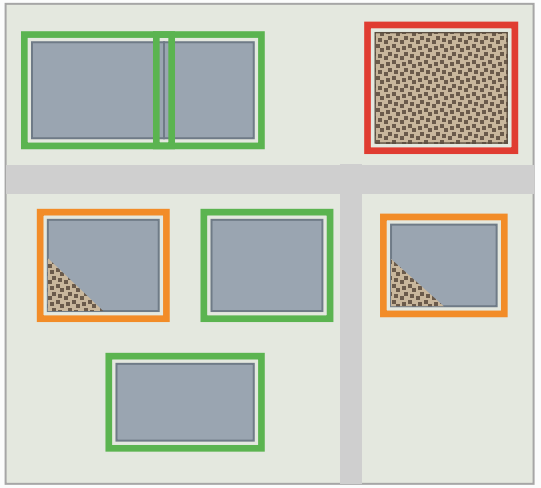
+ Precise localization and no crop size to choose

– Adjacent buildings **merge** into one region

– No building-level reasoning without post-processing

– Pixel annotations are **costly**

Detection and instance segmentation



+ Reasons at the **building** level

+ Footprints (e.g. OSM) give building locations

+ Boxes, polygons, masks, or even points

+ Panoptic variants add roads, bridges, …

– Masks cost more than boxes

Instance-based formulations reason where decisions are made: at the building level.

Each formulation implies its own data, methods and scores

The intended use case should drive the choice of formulation.

Scientific benchmark: methodological diversity, fine-grained annotations.

Operational benchmark or service selection: practical relevance of the extracted information.

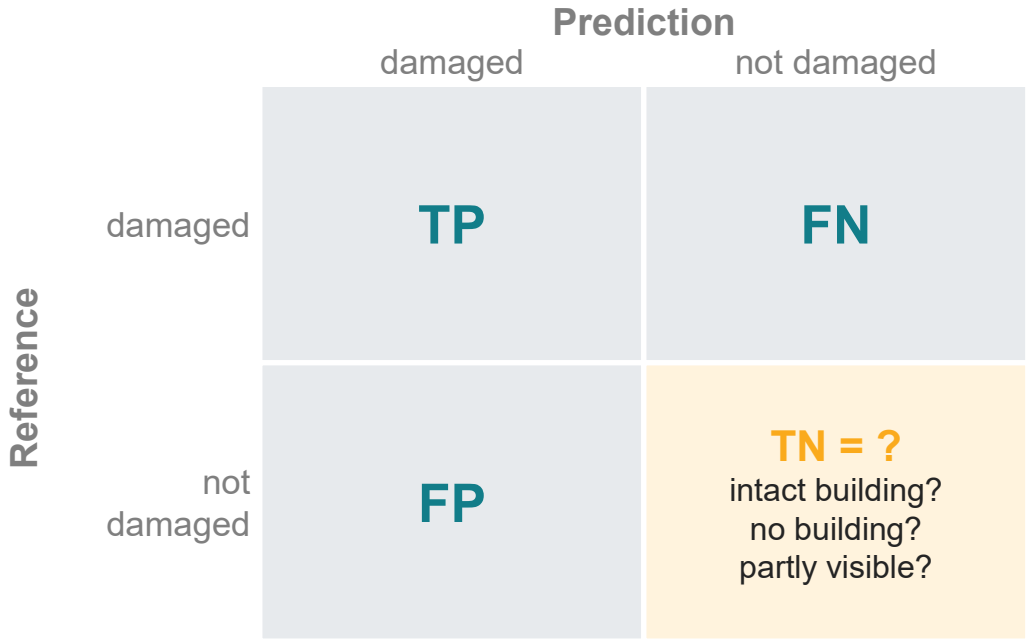
Formulation	Semantic level	Output	Annotation cost	Building-level reasoning	Localization
Binary classification	Image crop	Damaged / not damaged	Low	No	Implicit
Multi-class classification	Image crop	One damage class per crop	Low	No	Implicit
Semantic segmentation	Pixel	One class per pixel	High	No: instances merge	Precise
Detection / instance segmentation	Building instance	Shape + damage class per building	Medium (boxes, points) to high (masks)	Yes	Good to precise

The formulation changes the score, and the ranking

Property

- **Precision**, **recall** and **F1** depend on which class is called **positive**
- swapping damaged ↔ intact can change the **ranking** of methods.
- Merging ‘no building’ with ‘undamaged’ changes the confusion matrix itself.

Where does ‘no building’ go?



Agree on the task before comparing scores.

How to evaluate and rank BDA methods properly → **M. Van Droogenbroeck, 14:30**

Our proposal: predict damage per building instance

Idea

Output: every building instance, localized (box, polygon, mask or point) and assigned a damage class.

- Reasons at the **building level** and reuse existing **footprint** data.
- Balances **localization**, **annotation feasibility** and **operational interpretability**.

Instance-based: the prediction target of our task card.

BDA task card

Input data	?
Prediction target	**Every building instance, localized, with a damage class**
Reference labels	?
Evaluation	→ *M. Van Droogenbroeck, 14:30*

Table of content

JRC Workshop · Day 2 · 11 November	
09:00	**The three pillars of the BDA task** A. Cioppa
09:40	BDA: data quality A. Cioppa
10:30	Deep learning for BDA V. Blua
14:30	Performance-based ranking of BDA methods M. Van Droogenbroeck
Day 3	Hands-on: algorithms on VHR images V. Blua



Why revisit BDA?

- A decade of research, no shared foundation
- The three pillars
- The BDA task card

What should we predict?

- One task, many names
- Four candidate formulations
- Our proposal: building instances

From which data?

- Modality, time and geography
- Generalization and data splits
- Existing datasets

Three axes describe BDA data

Describe every dataset along three axes.

These axes are **intrinsic to the data** and independent of the task formulation: any formulation can be instantiated on any combination.

1 Modality

Which sensor?

Optical (Sentinel-2, VHR) · SAR · UAV · LiDAR · social media · building maps



2 Temporality

Post-event only, or pre + post?

Single-temporal analysis vs multi-temporal change detection



3 Geography and context

Which regions, building styles, hazards?

Architecture, urban density, materials, environment, disaster type



What each modality can, and cannot, reveal

No modality sees everything; optical and SAR remain the main modalities in datasets.



Optical: Sentinel-2

+ Free, 10 m, revisit of ~5 days

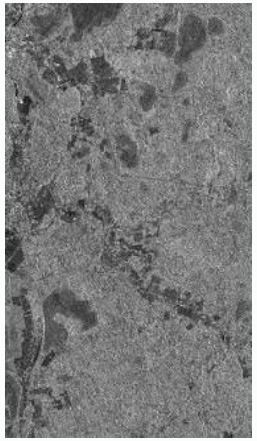
– Clouds; limited detail at 10 m



Optical: VHR (down to 0.3 m)

+ Individual buildings and heights visible

– Commercial tasking; oblique views; clouds



SAR (Sentinel-1, TerraSAR-X, …)

+ All-weather (e.g. tropical storms), day and night

– Speckle, distortions; Sentinel-1 alone insufficient for single buildings



UAV / aerial

+ cm resolution, oblique façade views, video, LiDAR

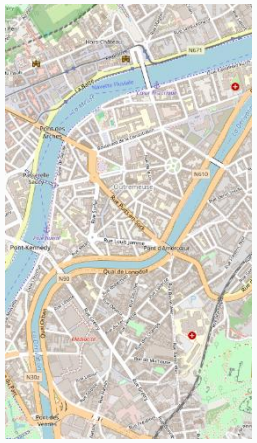
– Battery, weather, airspace rules, alignment errors



Social media

+ Ground-level views, very fast

– Must be geolocated; opportunistic



Building footprints (OSM, cadastre)

+ Locate buildings known before the event

– Volunteer-made: incomplete or outdated

Seen from above, partial damage is often invisible

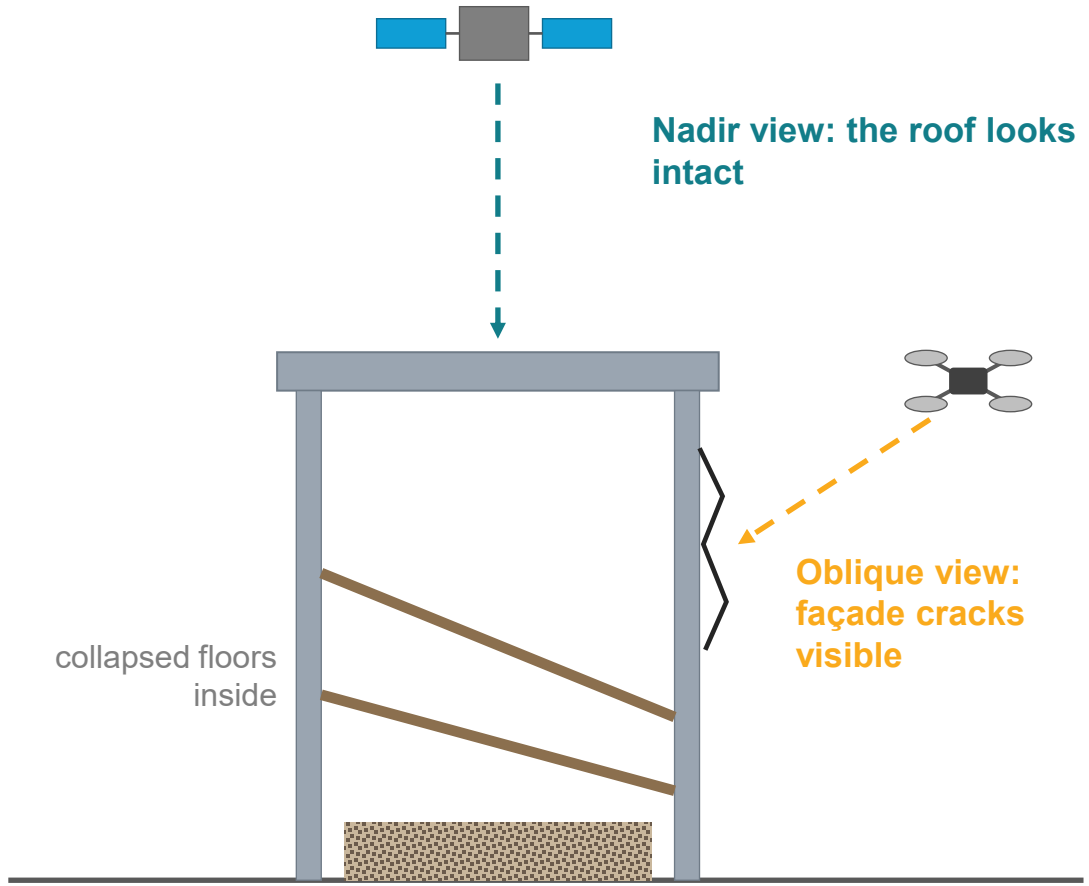
Limitations

Optical satellites reliably capture **collapsed** or **washed-away** buildings.

Intermediate damage *(façade cracks or internal collapse)*, is often hidden by the roof.

UAVs see façades, but coverage, regulations and alignment limit their use.

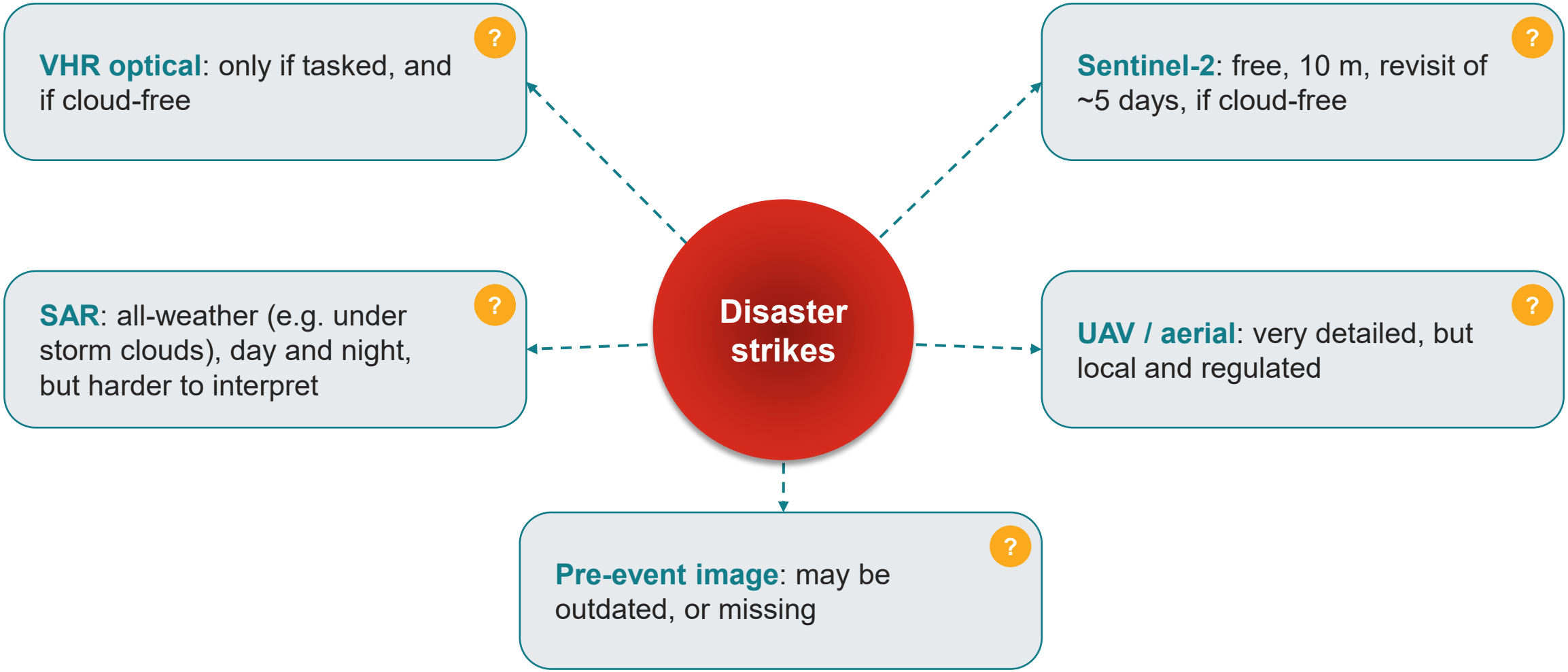
What cannot be seen from above cannot be detected from above.



Adriano et al., Remote Sensing 2019

In an emergency, you do not choose your data

The task must not assume a sensor, nor rely on pre-event imagery.

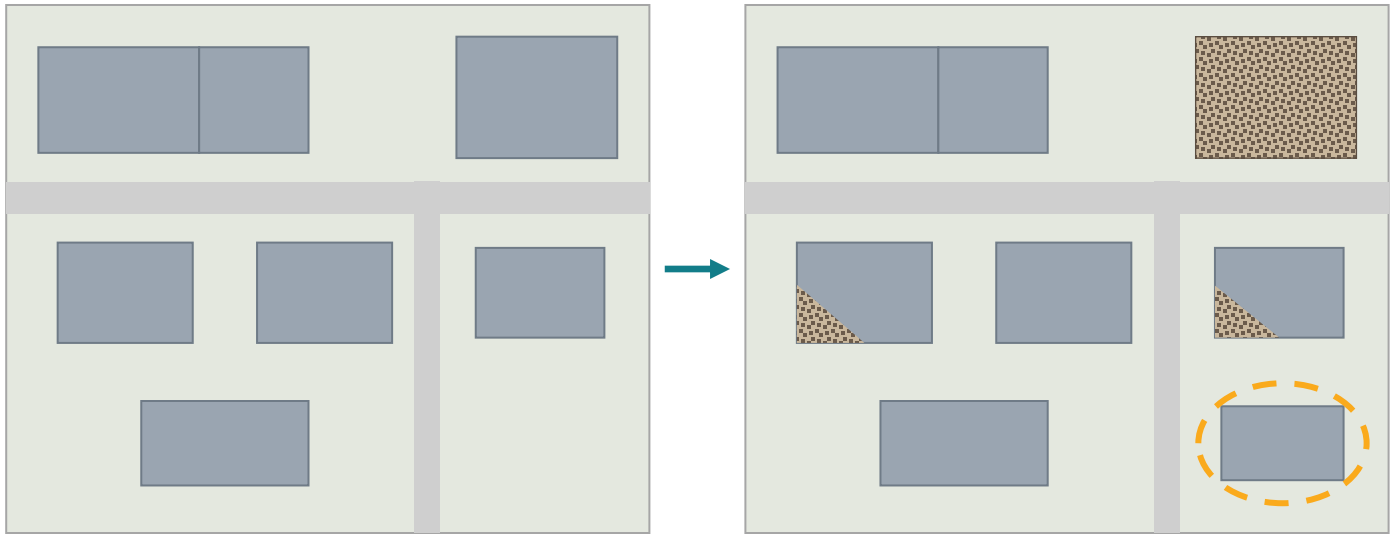


Pre-event imagery helps, but is never guaranteed

Problem statement

In some regions, recent pre-event imagery may be **missing** or **outdated**.

New construction looks like **change**; different **viewing angles** between pre and post create false differences.



Principle

Operational BDA should rely primarily on **post-event** observations; pre-event data should be considered as **optional** supplementary information.

Models struggle to generalize to new regions

Results

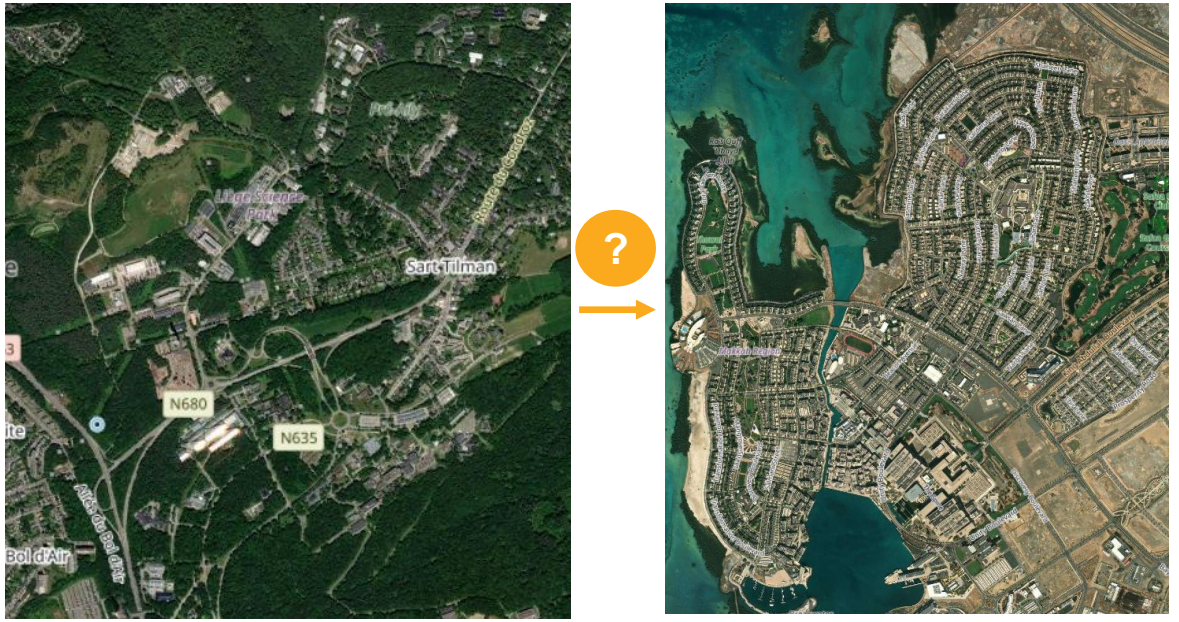
No reviewed approach shows strong robustness **across regions**.

Leave-one-city-out is a step forward, but stays within one homogeneous context.

Many studies still train and test on **one event, one region**.

Benchmarks must test what deployment requires: **unseen regions.**

Training region **New region**



Different architecture, layouts, materials and environment: damage **looks** different.

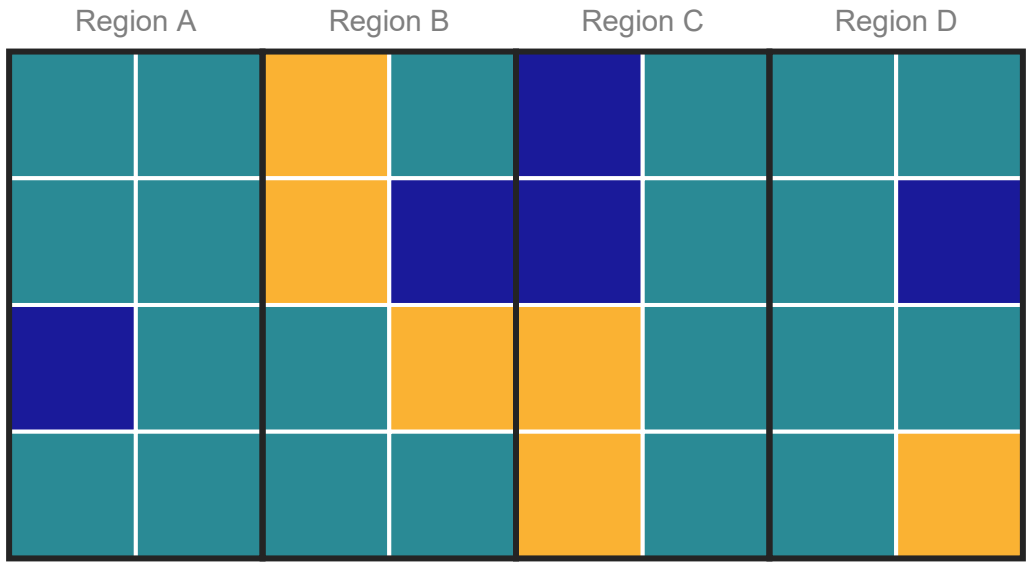
Gu et al., Buildings 2024; Russo et al., IEEE JSTARS 2026

Random splits leak, geographic splits do not

Principle

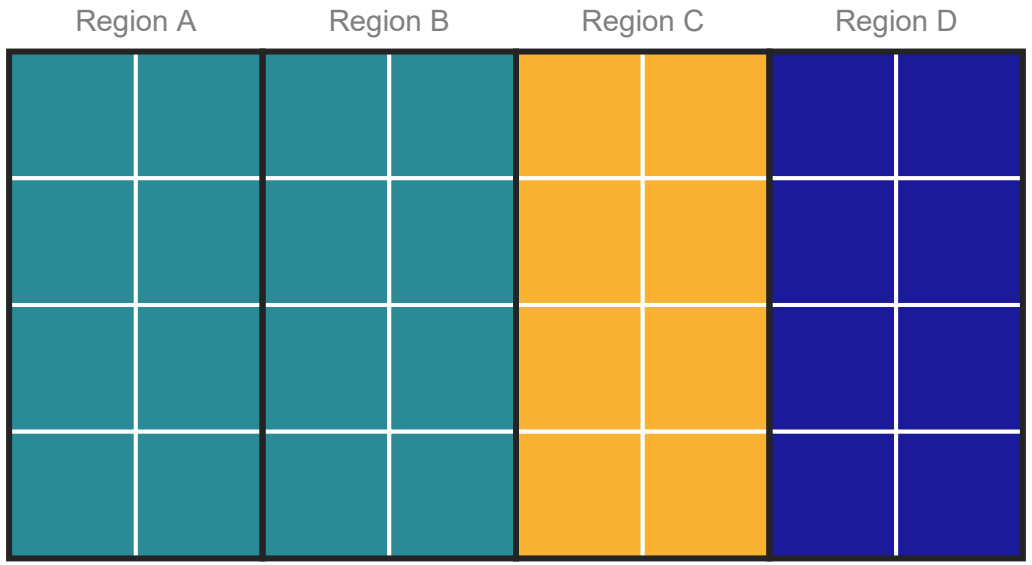
Geographically disjoint splits is a minimal requirement for training and evaluating AI models.

Random split



Neighboring tiles of the same city end up in training **and** test: optimistic scores (**spatial leakage**).

Geographically disjoint split



Each region–disaster pair should appear in **one** split only.

Training Validation Test

Data-sharing setups and trust → **M. Van Droogenbroeck, 14:30**

The dataset landscape is fragmented

Dataset (year)	Task	Modality	Temporality	Region	Buildings	Annotation
ABCD (2017)	Binary	Aerial (PASCO)	Pre + post	Japan	22,171	Image crop
Nguyen et al. (2017)	Multi-class	Social media	Post	4 regions	25,758	Image-level (3)
Cheng et al. (2021)	Multi-class	UAV	Post	Caribbean	6,440	Instance masks (2)
Xue et al. (2024)	Multi-class	Street view + struct.	Post	Florida, USA	2,472	Building-level (3–5)
FloodNet (2021)	Class., seg., VQA	UAV	Post	Texas, USA	8,946	Pixel-level (9) + VQA
RescueNet (2023)	Class., semantic seg.	UAV	Post	Florida, USA	10,903	Pixel masks + image (10)
Ida-BD (2023)	Multi-class, seg.	VHR optical	Pre + post	Louisiana, USA	N/A	Instance polygons (4)
CRASAR-U-DROIDS (2024)	BDA	sUAS orthomosaic	Post	USA	21,716	Instance polygons (4–5)
Noto-BDA-MV (2025)	Binary	Optical + auxiliary	Pre + post	Japan	140,208	Instance polygons
xBD (2019)	**Multi-class**	**Optical satellite**	**Pre + post**	**Worldwide**	**850,736**	**Instance polygons (4)**
xFBD (2022)	Multi-class, inst. seg.	Optical satellite	Pre + post*	Palu, Indonesia	8,027	Instance polygons (4)
QuickQuakeBuildings (2024)	Binary	Optical + SAR	Post	Islahiye, Türkiye	4,029	Building-level binary
BDD (2021)	Semantic seg.	Optical + SAR	Pre + post	Multiple regions	N/A	Pixel masks (3)
BRIGHT (2025)	**Class., seg.**	**Optical + SAR**	**Pre + post**	**Worldwide**	**384,596**	**Instance polygons (4)**

xBD and BRIGHT: steps in the right direction



xBD (2019)

- **19 events** (11 major + 8 additional), 2011-2019: earthquakes, wildfires, tornadoes, floods, volcanoes
- **Pre + post** optical imagery (Maxar Open Data)
- **Building polygons**, 4 damage classes, two rounds of quality control
- **Problem:** unclear whether regions overlap across splits



BRIGHT (2025)

- **14 events, 23 regions**, 5 natural + 2 human-made disaster types
- **VHR optical (pre) + SAR (post)**: all-weather
- **Building-level** annotations; open-source
- **Problem:** commercial VHR SAR raises availability and cost questions in practical development

Multi-event, multi-region data: the right direction. What models achieve on them → **V. Blua, 10:30**

Part 1 in one slide

Summary

BDA is defined inconsistently and formulations are **not interchangeable**.

Prediction target: every **building instance** with a damage class.

Inputs: **any available sensor**, post-event first, pre-event optional.

Evaluation on **unseen regions**, with geographically disjoint splits.

Next: can we trust the data, and the labels?

BDA task card

Input data	**Any available sensor · post-event first, pre-event optional · tested on unseen regions**
Prediction target	Every building instance, localised, with a damage class
Reference labels	?
Evaluation	→ *M. Van Droogenbroeck, 14:30*

Table of content

JRC Workshop · Day 2 · 11 November	
09:00	The three pillars of the BDA task A. Cioppa
09:40	**BDA: data quality** A. Cioppa
10:30	Deep learning for BDA V. Blua
14:30	Performance-based ranking of BDA methods M. Van Droogenbroeck
Day 3	Hands-on: algorithms on VHR images V. Blua



Can we trust the inputs?

- Access to very-high-resolution imagery
- Geometry and co-registration
- Building footprints

Can we trust the reference?

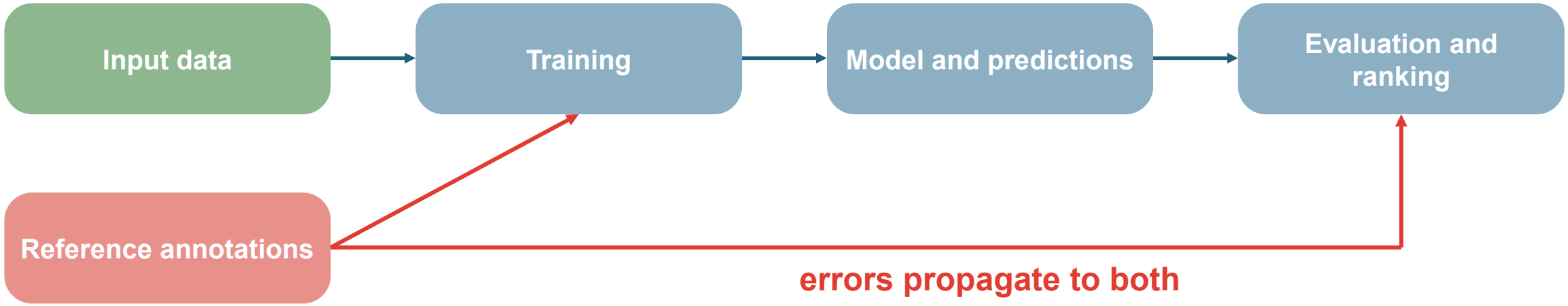
- Annotation typologies
- Missing and ambiguous classes
- Annotation under time pressure
- Towards an unambiguous typology

Recap: data quality concerns inputs and reference

Reminder

Part 1: what to predict (building instances), from which data (any sensor, post-event first).

Part 2: are the **inputs** and the **reference annotations** good enough to learn from, and to evaluate against?



Very-high-resolution imagery is not always accessible

Limitations

VHR images are not openly available to the scientific community.

Commercial tasking: satellites acquire only when commissioned, at a cost, with no guarantee of sufficient quality.

Clouds make optical images unusable. Waiting for the next pass is not acceptable in an emergency.

Two scenarios

Scientific exploration

Homogeneous data: one modality, fixed resolution, since results depend on the resolution used.

Deployment testing

Heterogeneous data that mimic real-world constraints.

Why AI tech sovereignty matters for Europe

Definition

"Tech sovereignty is Europe's ability to act independently in the digital world by developing and controlling key technologies, data, and infrastructure, while reducing reliance on non-EU providers."

> 80%

of key digital products, services, infrastructure and IP: the EU relies on non-EU countries

EC, citing CERRE (2022)

What BDA needs	What the EU plans
Compute & cloud train and run models fast, at scale	**Cloud and AI Development Act**: "at least tripling the EU's data center capacity within the next 5–7 years", one EU sovereignty framework with 4 assurance levels for public sector bodies
Models & adoption AI products for emergency services	**AI Continent Action Plan**: 5 key areas: computing infrastructure - **data** - skills - adoption - simplification **Apply AI**: a "buy European" approach, "particularly for the public sector", with open-source AI. Sectors include "climate and environment" and "defence, security and space"
Data imagery, footprints, labels	**Data Union Strategy** and **Data Act** → **next slide**

Over-reliance on non-EU cloud is "a significant risk to Europe's digital autonomy and resilience".

For rapid building damage assessment, resilience cannot fail.

Sources: European Commission, digital-strategy.ec.europa.eu: Tech sovereignty · AI Continent Action Plan · Apply AI · Cloud and AI Development Act.

AI sovereignty starts with data

"the EU must be able to fuel AI with high-quality data" *Data Union Strategy*

Access to data for AI

Data Union Strategy

- Data labs "pool both private and public resources to make high-quality sectoral data available"
- Common European data spaces: around €100 million of ongoing EU investment

For BDA: pooled post-event imagery and labels, not fragmented datasets.

Emergency access to data

Data Act

Public sector bodies can "request data necessary to help them respond quickly and securely to a public emergency"

- Easy switching between cloud providers

For BDA: a disaster is a public emergency: rapid mapping may request privately held data, such as commercial VHR.

Data sovereignty

Data Union Strategy

- Guidelines on "fair treatment of EU data abroad"; a toolbox to counter data leakage
- "Sovereignty requires openness to trusted partners […] on terms that are fair, secure, and consistent with EU values and interests."

For BDA: international imagery and models stay welcome, but the EU need their own secured data.

AI Act: for high-risk AI systems, Europe requires "high-quality of the datasets feeding the system", a "high level of robustness, cybersecurity and accuracy" and "appropriate human oversight measures".

Data quality is a sovereignty issue: Europe's AI is only as trustworthy as its data.

Sources: European Commission, digital-strategy.ec.europa.eu: Data Union Strategy · Data Act · AI Act.

Geometry problem: off-nadir views and co-registration

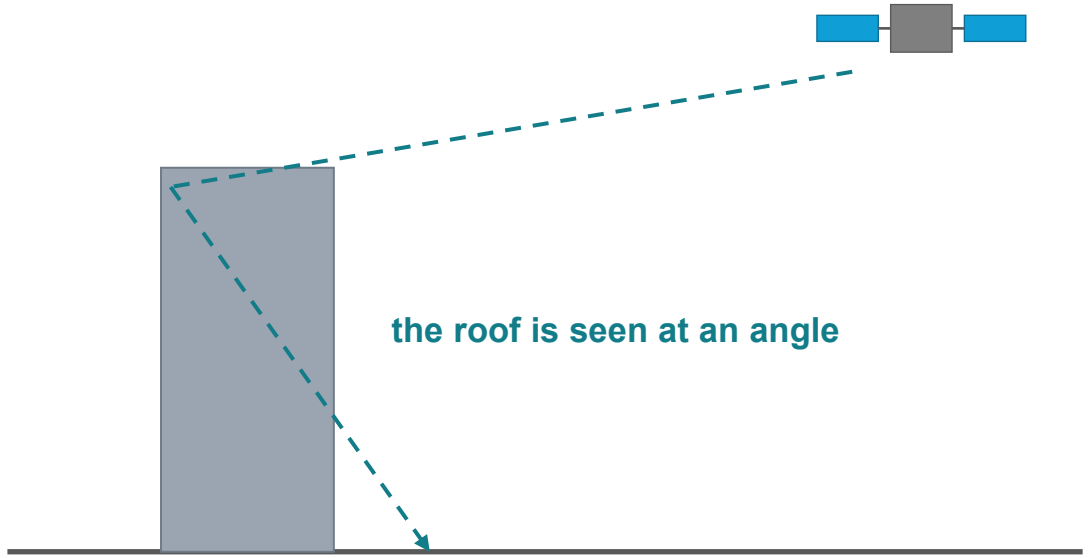
Problem statement

VHR images are captured **at an angle**: roofs are displaced from footprints.

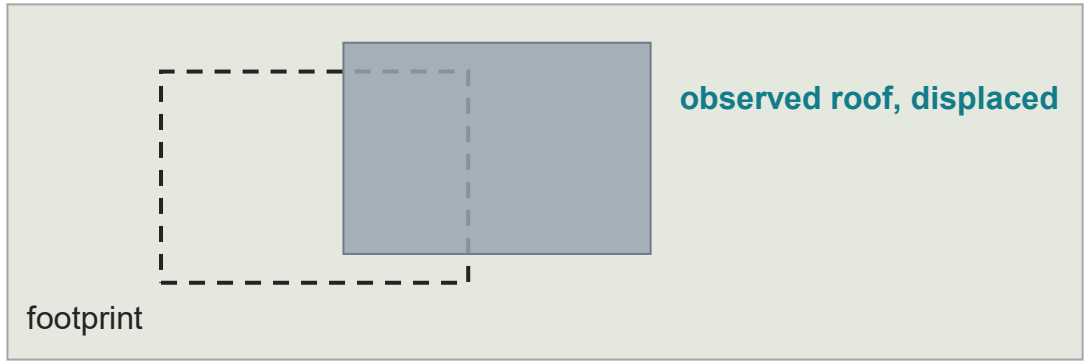
Geolocating damaged buildings is therefore difficult.

Co-registering **SAR with footprints** relies on visible wall bases: it assumes buildings are still **upright**

Higher resolution does not automatically mean more reliable annotations.



Top view



Sun et al., 2024

Building footprints are useful, but imperfect

Overview

Footprints locate buildings **known before the event** and turn pixel problems into building-level ones.

OSM is volunteer-made: **completeness**, **accuracy** and **update frequency** vary across regions.

Footprint errors propagate into datasets and predictions.

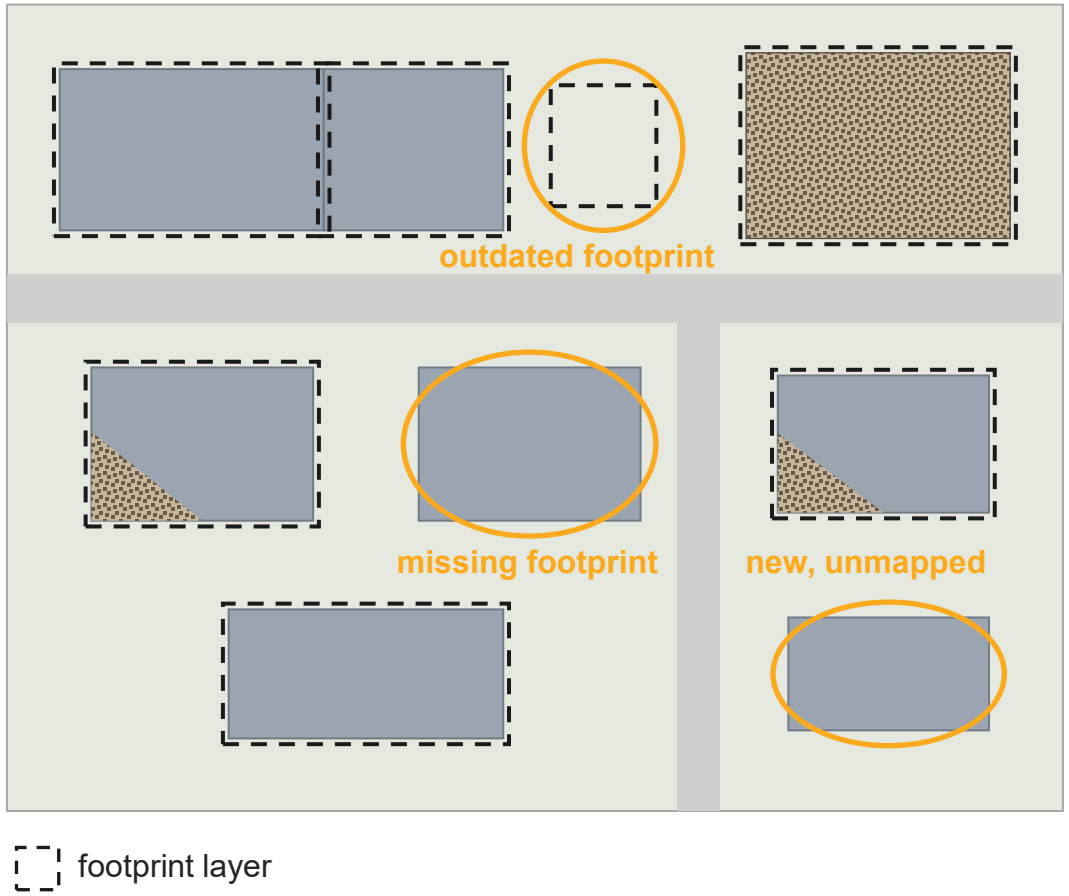


Table of content

JRC Workshop · Day 2 · 11 November	
09:00	The three pillars of the BDA task A. Cioppa
09:40	**BDA: data quality** A. Cioppa
10:30	Deep learning for BDA V. Blua
14:30	Performance-based ranking of BDA methods M. Van Droogenbroeck
Day 3	Hands-on: algorithms on VHR images V. Blua



Can we trust the inputs?

- Access to very-high-resolution imagery
- Geometry and co-registration
- Building footprints

Can we trust the reference?

- Annotation typologies
- Missing and ambiguous classes
- Annotation under time pressure
- Towards an unambiguous typology

Different typologies for the same damage

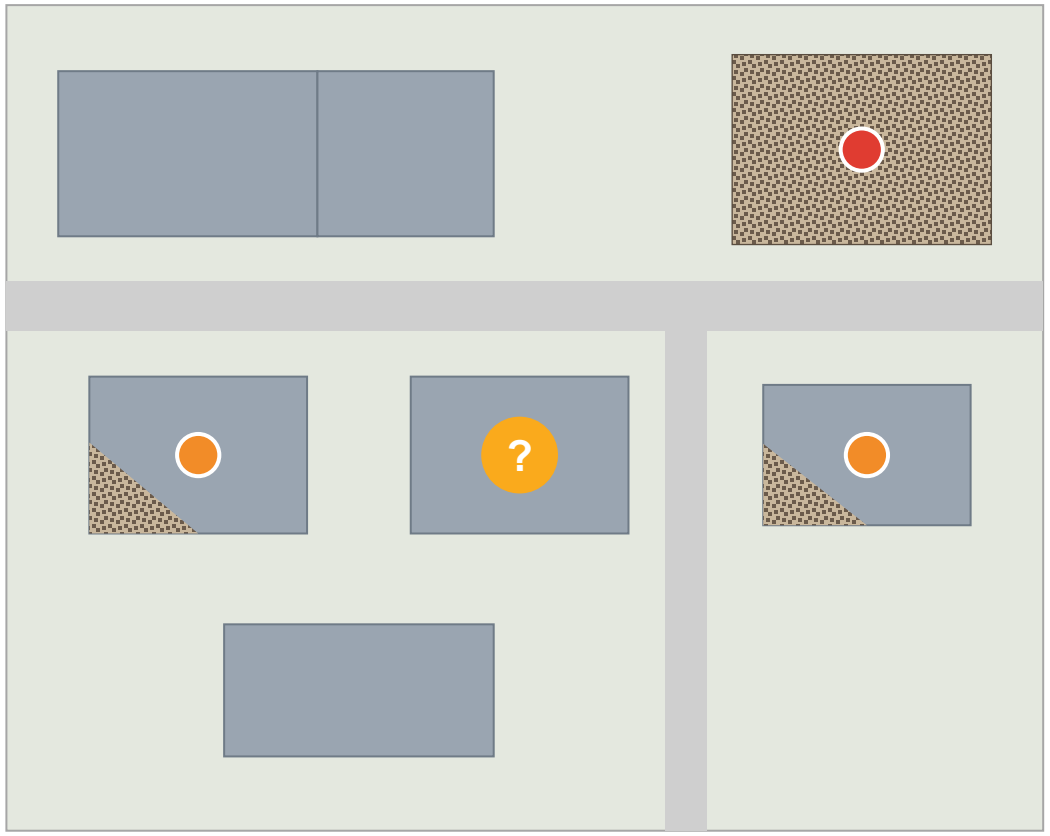
Limitations

Same names may lead to different meanings and different names the same ones.

Labels do not transfer across datasets. Most benchmarks then collapse classes to two.

IWG-SEM guidelines	CEMS ODM	xBD	BRIGHT
4 levels, with definitions and examples	On-Demand Mapping: points, manual visual interpretation	building polygons, 4 classes	building level, 3 classes + background
Destroyed	Destroyed	Destroyed	Destroyed
Damaged	Damaged	Major damage / Minor damage	Damaged
Possibly damaged	Possibly damaged		
No visible damage		No damage	Intact
			Background

An unlabeled building can mean three things



Destroyed Damaged *point annotations, CEMS-like*

No label on a building means…?

Intact
assessed, no visible damage

Background
not a building

Never assessed
outside the area, under clouds, or no time

“Possibly damaged” is uncertainty, not a damage state

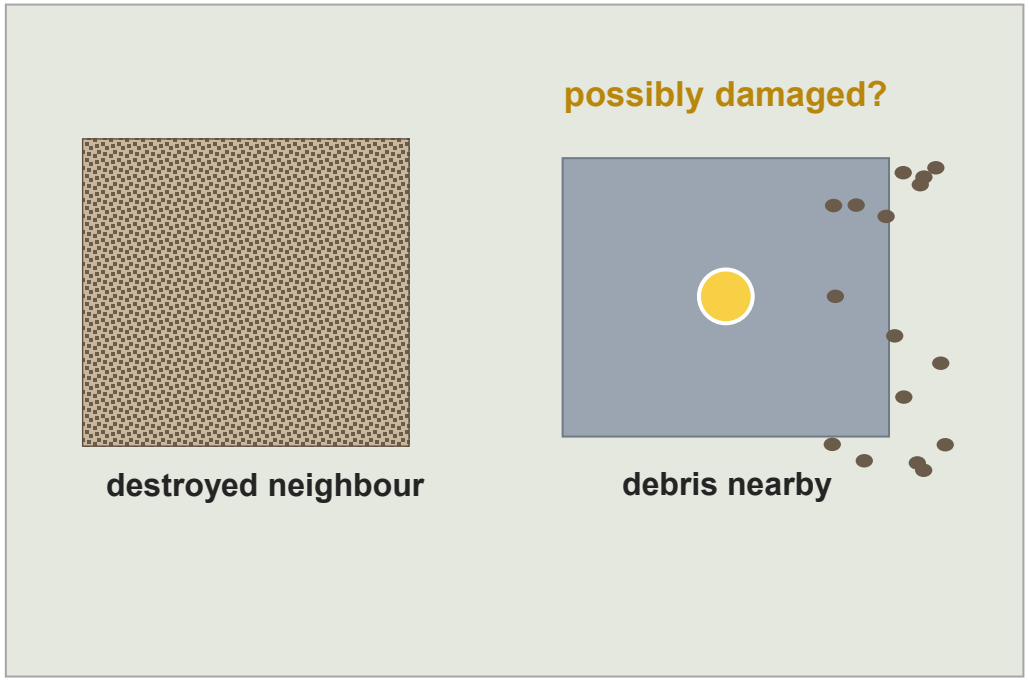
Properties

Possibly damaged is assigned when damage is **suspected**: debris nearby, or a damaged neighbor.

It reflects **annotator uncertainty** rather than a physical state and is open to interpretation.

Divergence: the BRIGHT authors reclassify it as **intact**.

Do not treat it on an equal footing with the other classes.



Annotations made under time pressure are imperfect

Class errors

Debris or a damaged neighbor lead to a ‘possibly damaged’ label, or a wrong class.

Geolocation and alignment

Acquisition angles make damaged buildings hard to locate precisely.

Heterogeneous shapes

Masks, boxes, segments, convex shapes, or a single point per building.

Missed areas

Under time pressure, some parts are simply not annotated.

Hand-made references are best estimates, not ground truth.

Towards an unambiguous typology

Proposed classes

- **Destroyed**
- **Damaged + score**
- **Potentially damaged: visible damage or debris in the surroundings**
- **No visible damage**
- **Background**
- **Cloud mask**

Idea

Footprints (OSM, cadastre) localise buildings, so **background** becomes an explicit class.

Potentially damaged: debris or a damaged neighbour; damage may hide under an intact roof.

Simplest annotation: one **point** per building on the footprint → instance-based segmentation.

Common **cloud masks**: clouds are unlabelled areas, not background.

Ideally, a **damage rate** (0–100 %) with a **confidence** score.

Re-annotation has a cost, but an unambiguous label is always preferable.

Recommendations for BDA datasets

Summary

Trustworthy data first

Define the task and each class precisely.

Align annotations with the application and evaluation.

Separate intact, background and not assessed and handle ‘potentially damaged’ explicitly.

Provide common building footprints and cloud masks.

Split geographically: each region–disaster pair in one split only.

BDA task card

Input data	Any available sensor - post-event first, pre-event optional - tested on unseen regions
Prediction target	Every building instance, localised, with a damage class
Reference labels	**Unambiguous classes - explicit background and 'not assessed’ - documented quality**
Evaluation	*→ M. Van Droogenbroeck, 14:30*

Conclusion and further reading

Summary

Define the task first: formulations are not interchangeable. We favor an instance-based, sensor-agnostic definition.

Know the data: modality, timing and geography control what can be learned and what can be claimed.

Question the literature: typologies and annotation practice affect both training and evaluation.

Thanks, questions?

Revisiting the Task of Building Damage Assessment by Satellite Imaging

A. Cioppa, V. Blua, S. Piérard, P. Ceccato, M. Van Droogenbroeck

arXiv:2610.13857 (2026)

Coming up today: methods and evaluation

BDA task card: complete

Input data	Any available sensor - post-event first, pre-event optional - tested on unseen regions
Prediction target	Every building instance, localised, with a damage class
Reference labels	Unambiguous classes - explicit background and 'not assessed' - documented quality
Evaluation	→ *M. Van Droogenbroeck, 14:30*

10:30 · Deep learning for BDA

V. Blua

From task-specific networks to Earth-observation foundation models; generalisation to unseen events and regions. Day 3: hands-on on VHR images.

14:30 · Performance-based ranking

M. Van Droogenbroeck

Why a single score is fragile; rankings and multiple scores (the Tile); data-sharing setups, trust and golden samples.