

Post-Earthquake Candidate Building-Loss Screening from Bitemporal Google Earth Imagery: An Object-Level Framework for the 2023 Türkiye Earthquakes

perihan karaköse

pkarakose@bartin.edu.tr

Bartın Üniversitesi <https://orcid.org/0000-0002-8894-6997>

Research Article

Keywords: earthquake damage assessment, building detection, remote sensing, artificial intelligence, decision support

Posted Date: August 31st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10670554/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Post-Earthquake Candidate Building-Loss Screening from Bitemporal Google Earth Imagery: An Object-Level Framework for the 2023 Türkiye Earthquakes

Perihan Karakose^{1*}

^{1*}Department of Electronics and Automation, Bartın University,
Bartın, 74100, Turkey.

Corresponding author(s). E-mail(s): pkarakose@bartin.edu.tr;

Abstract

Rapid and reliable identification of building losses after major earthquakes is essential for post-disaster assessment and decision support; however, many existing approaches depend on accurately coregistered imagery, building footprints, geographic information system layers, or other auxiliary geospatial datasets that may not be consistently available across large affected regions. This study proposes an object-level bitemporal framework for automatically screening candidate building losses from historical Google Earth imagery acquired before and after the 6 February 2023 Türkiye earthquakes. A hybrid building-detection dataset was constructed by combining the Satellite-II Version 1 dataset with manually verified Google Earth images from earthquake-affected areas. Six representative object detectors—Faster R-CNN, RetinaNet, SSD, YOLOv10n, YOLOv11n, and YOLOv12n—were evaluated under a harmonized experimental protocol. YOLOv12n provided the most favorable accuracy–efficiency trade-off and was subsequently retrained using the complete hybrid dataset. On an independent Google Earth test set, the final model achieved 72.19% precision, 33.26% recall, and a 45.54% F1-score, indicating substantially improved precision after full-data training. For the temporal analysis, corresponding 2022 and 2024 images were spatially aligned using stable non-building references, including road and background structures, followed by object-level one-to-one matching based on geometric correspondence. A total of 145 bitemporal image pairs were processed. The detector identified 2,872 buildings in 2022 and 2,039 buildings in 2024; 842 buildings were successfully matched across acquisition dates, whereas

2,030 unmatched detections were retained as candidate building losses, yielding an initial automatic candidate-loss rate of 70.68%. The results demonstrate the potential of publicly accessible historical Google Earth imagery for scalable preliminary post-earthquake building-loss screening without mandatory reliance on building footprints, LiDAR, or cadastral data. The proposed framework is intended as a decision-support tool for prioritizing locations requiring detailed expert assessment rather than as a stand-alone system for definitive damage classification.

Keywords: earthquake damage assessment, building detection, remote sensing, artificial intelligence, decision support

1 Introduction

Earthquakes can cause extensive structural destruction within seconds and disrupt transportation, communication, health care, energy supply, and other critical urban services. Immediately after a major earthquake, emergency-management authorities require rapid and spatially consistent information on the location of collapsed or demolished buildings to prioritise search-and-rescue operations, allocate limited resources, identify accessible routes, and support recovery planning. Conventional engineering surveys provide detailed structural evidence, but their implementation over large urban regions is labour-intensive, time-consuming, and often restricted by unsafe or inaccessible conditions. Remote sensing and computer vision therefore provide an important complementary pathway for large-area post-disaster assessment.

Ground-level imagery can reveal cracks, spalling, debris, and façade collapse in considerable detail. Liu et al. [1] developed a lightweight YOLOv5-based model for detecting several damage manifestations from terrestrial images. Although such systems are suitable for mobile and field-based inspection, their dependence on physical access, suitable viewpoints, and dense image acquisition limits their scalability in disasters affecting extensive regions. Satellite and aerial imagery overcome part of this limitation by offering synoptic coverage and enabling the analysis of the same urban area at multiple dates.

A large body of research has addressed post-disaster building assessment using very-high-resolution satellite imagery, orthophotos, unmanned aerial vehicle imagery, Synthetic Aperture Radar, LiDAR, and geographic information system data. Studies have formulated the task as image classification [2–6], semantic or instance segmentation [7–9], object detection [10, 11], anomaly detection [12, 13], and bitemporal change detection [14–20]. Public datasets such as xBD have substantially accelerated this field by providing pre-event and post-event imagery with building polygons and ordinal damage labels [21]. Nevertheless, the relationship between high benchmark performance and reliable deployment remains uncertain because disaster imagery varies in acquisition date, viewing angle, ground sampling distance, illumination, season, and georegistration quality.

Bitemporal analysis is particularly relevant to post-earthquake building-loss assessment because a destroyed or removed building may be easier to infer from its disappearance or major structural change than from post-event appearance alone. Classical change-detection methods compare spectral, textural, geometric, or height-related features between dates [14, 18–20, 22, 23]. More recent approaches integrate pre-event and post-event image features through convolutional or transformer-based networks [9, 13, 15–17, 24]. However, many of these methods require accurately coregistered image pairs and pixel-level change masks. Google Earth imagery presents a realistic but challenging setting because images from different years may differ in scale, viewing direction, shadows, seasonal conditions, compression, and local geometric displacement.

This study adopts a detection-first bitemporal strategy. The public Satellite-II Version 1 building-detection dataset [25] is augmented with 150 Google Earth images collected from regions affected by the 6 February 2023 Türkiye earthquakes. The additional images are automatically annotated and subsequently inspected and corrected manually. Six representative object detectors—SSD, Faster R-CNN, RetinaNet, YOLOv10, YOLOv11, and YOLOv12—are screened under harmonised data partitions, a limited computational budget, and common evaluation metrics while retaining architecture-specific training requirements. The detector providing the most favourable accuracy-efficiency trade-off is then retrained on the complete hybrid dataset and applied independently to Google Earth imagery acquired in 2022 and 2024.

The resulting pre-event and post-event building inventories are compared after scale normalisation and spatial alignment. Buildings detected in 2022 but lacking reliable counterparts in 2024 are treated as candidate earthquake-related building losses rather than definitive earthquake-induced demolitions. This cautious interpretation accounts for alternative explanations, including planned demolition, reconstruction, image-quality differences, shadows, occlusion, and missed detections. The proposed workflow therefore separates robust building localisation from object-level temporal reasoning and requires manual verification of unmatched building candidates.

The principal contributions are: (1) construction of a quality-controlled hybrid building-detection dataset combining Satellite-II Version 1 with earthquake-region Google Earth imagery; (2) staged cross-architecture comparison of representative one-stage and two-stage detectors under common data partitions and evaluation metrics; (3) full-dataset retraining of the detector selected according to a predefined accuracy-efficiency criterion; (4) generation of temporally comparable building inventories from independent 2022 and 2024 Google Earth imagery; and (5) object-level matching and manual verification of candidate earthquake-related building losses.

1.1 Related Work

Research on automated post-disaster building assessment has progressed from conventional post-event image interpretation toward increasingly sophisticated classification, segmentation, object detection, anomaly detection, and bitemporal change-analysis approaches. Existing studies differ considerably in terms of image source, temporal

configuration, auxiliary geospatial information, annotation requirements, and detection architecture. Accordingly, the existing literature is reviewed under four main themes: post-disaster damage recognition, pre- and post-event building damage assessment, building change detection and demolition mapping, and object detector families for building localisation.

1.1.1 Post-Disaster Damage Recognition from Ground-Level and Post-Event Images

Early computer-vision approaches commonly relied on post-event appearance. Liu et al. [1] used terrestrial imagery and a lightweight YOLOv5 model to recognise cracks, spalling, debris, and collapse. Ji et al. [4] classified collapsed and non-collapsed buildings from post-earthquake QuickBird image chips, while later studies investigated CNN features, transfer learning, and multi-class post-disaster image classification [5, 6, 26]. Zhan et al. [8] used a modified Mask R-CNN for instance-level extraction of damaged buildings, and Li et al. [10] investigated SSD for damaged-building detection from aerial imagery. These studies demonstrate the value of learned visual features, but post-event-only recognition may confuse rubble, bare ground, construction sites, shadows, and roof material changes.

Weakly supervised and anomaly-detection strategies have been proposed to reduce dependence on pixel-level or object-level damage annotations. Ali et al. [12] developed a weakly supervised framework that localises destruction from post-disaster satellite imagery using only image-level labels. Tilon et al. [13] proposed an anomaly-detecting generative adversarial network trained on undamaged pre-disaster building imagery to identify anomalous post-disaster building appearance without requiring explicit damage annotations. Although these methods reduce annotation effort, their performance may degrade when transferred across different geographic regions, sensors, or imaging conditions.

1.1.2 Pre and Post Event Building Damage Assessment

Pre-event and post-event imagery provides explicit temporal evidence. Kalantar et al. [9] compared CNN architectures using pre- and post-event orthophotos and showed that combining temporal information can improve earthquake-damage recognition. Ji et al. [6] fine-tuned a pretrained VGG model on bitemporal satellite imagery and reported that augmentation improved collapsed-building recognition. Xia et al. [16] proposed BDANet for the 2023 Türkiye earthquake, using pre-event imagery to delineate buildings and fused pre/post features to assess damage.

Object-level and footprint-assisted methods have also been developed. Ghaffarian et al. [7] used OpenStreetMap building data to generate training masks and update a post-disaster building database from multi-temporal WorldView-2 imagery. Braik and Koliou [27] integrated georeferenced satellite imagery, GIS building information, image-chip extraction, and CNN classification for large-scale mapping. Döşâ et al. [15] combined building footprints with pre- and post-earthquake VHR imagery from Antakya and compared SegNet, PSPNet, and ResUNet. Xie et al. [17] proposed OCD-BDA using Google Earth imagery from Hatay; YOLOv8 was used to localise

buildings, corresponding pre/post object pairs were generated, and a classification network assessed object-level change. This study is methodologically close to the present work because it addresses Google Earth inconsistencies and separates localisation from change assessment.

The xBD dataset [21] enabled systematic learning from pre/post disaster imagery at scale. It contains building polygons and ordinal damage labels from multiple disasters and regions. Subsequent methods have used xBD for localisation, damage classification, transfer learning, and anomaly detection [12, 13, 24, 28]. However, xBD contains heterogeneous disaster types and image conditions, and performance on its curated pairs does not guarantee robust transfer to independently collected Google Earth images.

1.1.3 Building Change Detection and Demolition Mapping

Building change detection is broader than disaster assessment and includes new construction, demolition, and map updating. Su et al. [18] combined bitemporal aerial images, digital surface models, and an old building map and reported high recall for both new and demolished buildings. Khodaverdi Zahraee and Rastiveis [19] segmented pre/post WorldView-2 imagery, extracted object-level texture differences, and used an artificial neural network to identify changed and unchanged buildings after the Haiti earthquake. Earlier work also exploited height differences from pre- and post-seismic stereo imagery to detect collapsed structures [20]. These studies show that object existence and geometric change are informative cues for demolition.

More recent earthquake-specific change-detection studies use deep feature fusion. Xie et al. [17] explicitly avoided strict pixel-level differencing by forming object-level pre/post pairs. Dösa et al. [15] used building footprints as labels for semantic segmentation and then compared pre- and post-earthquake building conditions. Emerging work on short-interval earthquake change detection additionally addresses side-looking and feature misalignment through multi-scale interaction and offset calibration. Collectively, these studies support the use of object-level temporal reasoning when image pairs are imperfectly aligned.

1.1.4 Object Detector Families for Building Localisation

Object detection architectures exhibit substantial differences in both network design and operational characteristics. Faster R-CNN is a two-stage detector that first generates region proposals and subsequently performs object classification and bounding-box refinement, generally prioritizing localization accuracy over inference speed [29]. In contrast, SSD and RetinaNet are one-stage detectors. SSD performs dense predictions across multiple feature scales to improve multi-scale object detection [30], whereas RetinaNet introduces focal loss to mitigate the adverse effect of the overwhelming number of easy background samples during training [31]. The YOLO family also follows a one-stage detection paradigm, emphasizing end-to-end optimization and real-time inference efficiency [32–34]. Despite these advances, building detection in high-resolution remote sensing imagery remains challenging because buildings often appear as small, densely distributed, partially occluded objects with roof textures that are visually similar to roads, bare ground, and other man-made surfaces.

Previous studies on post-disaster building assessment have typically evaluated a single detection architecture or compared models within the same detector family [8–11, 35]. For example, Ilmak et al. [11] compared several YOLO variants for collapsed-building detection in Antakya using very-high-resolution satellite imagery. In contrast, the present study performs a systematic cross-architecture comparison involving SSD, Faster R-CNN, RetinaNet, YOLOv10, YOLOv11, and YOLOv12 using a common single-class building dataset, identical data partitions, and a harmonized evaluation protocol. This experimental design aims to isolate the influence of detector architecture by minimizing variations arising from differences in annotation strategy, geographic coverage, and evaluation data, thereby enabling a more objective assessment of detection accuracy and computational efficiency.

1.2 Research Gap and Contributions

As summarised in Table 1, previous studies have substantially advanced automated post-disaster building assessment through classification, segmentation, object detection, anomaly detection, footprint-assisted analysis, and bitemporal change detection. Nevertheless, four recurring gaps remain. First, many approaches infer damage solely from post-event appearance and may confuse structural destruction with debris, shadows, construction activity, roof-material variation, or imaging artefacts. Second, numerous bitemporal methods depend on accurately coregistered image pairs, building footprints, LiDAR, digital surface models, or manually prepared damage annotations that may not be consistently available across large earthquake-affected regions.

Third, many studies produce image-chip classes or pixel-level masks rather than independently detected building inventories that can be matched across acquisition dates. Fourth, detector comparisons are often restricted to a single architectural family or are conducted with different data partitions, image resolutions, training settings, and evaluation implementations. A controlled cross-architecture screening of representative one-stage and two-stage detectors, followed by full-data retraining of only the selected model, remains comparatively underexplored in this application domain.

Historical Google Earth imagery also introduces a specific methodological challenge. Images acquired at different dates may vary in scale, viewing angle, spatial resolution, illumination, seasonal conditions, compression, and local geometric alignment. These inconsistencies reduce the reliability of direct pixel-wise differencing and motivate an object-level temporal inventory approach in which building localisation and temporal matching are evaluated as separate stages.

To address these gaps, the study follows a staged experimental design. During the screening stage, SSD, Faster R-CNN, RetinaNet, YOLOv10, YOLOv11, and YOLOv12 are evaluated using identical train-validation-test image identities and common detection metrics, while architecture-specific input sizes and optimisation requirements are retained. To reduce spatial data leakage, images originating from the same or overlapping geographic sites should be assigned to only one data partition. The preliminary comparison is conducted under a limited computational budget rather than through exhaustive full-dataset training of every model.

Primary model ranking should be based on mAP@0.50:0.95. A predefined tolerance rule should be used when models obtain similar detection performance; within

that tolerance, inference latency, FPS, and training cost can be used to select the model providing the most favourable accuracy-efficiency trade-off. The selected detector is then retrained using the complete hybrid dataset. Detailed settings such as maximum epochs, early-stopping patience, batch size, optimiser, and learning-rate schedule should be reported in the Experimental Setup section rather than presented as research contributions.

In the temporal-analysis stage, the selected detector is applied independently to the 2022 and 2024 Google Earth images. The images must first be brought to a common spatial scale and aligned using an explicitly documented procedure. Building objects are then matched using complementary geometric evidence, such as buffered Intersection over Union, centroid distance, relative area consistency after scale normalisation, and neighbourhood agreement. Unmatched 2022 detections are retained as candidates and checked against the original 2024 imagery.

For scientific validation, candidate building losses should be reviewed independently by at least two assessors using predefined categories, such as demolished, unchanged or missed, reconstructed or footprint-altered, and uncertain. Disagreements should be resolved by consensus, and inter-rater agreement should be reported where feasible. This procedure is necessary because absence of a model detection is not, by itself, proof of earthquake-induced demolition.

The principal contributions of this study are as follows:

1. **Quality-controlled hybrid dataset:** Satellite-II Version 1 is combined with 150 Google Earth images from the 2023 Türkiye earthquake region. Automatically generated annotations are manually inspected and corrected before model development.
2. **Staged cross-architecture detector comparison:** SSD, Faster R-CNN, RetinaNet, YOLOv10, YOLOv11, and YOLOv12 are screened using common image partitions and evaluation metrics under a limited computational budget, while architecture-specific training requirements are retained.
3. **Predefined accuracy-efficiency model selection:** The final detector is selected using a prespecified rule centred on mAP@0.50:0.95 and, when performance differences are within a defined tolerance, computational measures such as latency, FPS, and training cost.
4. **Temporal building-inventory generation:** The selected detector is retrained on the complete hybrid dataset and applied independently to 2022 and 2024 Google Earth imagery to create comparable building inventories.
5. **Object-level candidate building-loss mapping:** After documented scale normalisation and spatial alignment, detected buildings are matched across dates. Unmatched pre-event buildings are treated as candidate earthquake-related building losses and are subjected to manual verification rather than being automatically labelled as earthquake-induced demolitions.

Overall, the proposed framework provides a computationally constrained and methodologically transparent workflow that separates detector development from temporal change analysis. It avoids attributing every unmatched building directly to

the earthquake and instead produces a manually verifiable inventory of candidate earthquake-related building losses from historical Google Earth imagery.

The remainder of this paper is organized as follows. Section 3 presents the theoretical background of the evaluated object detection architectures together with the performance metrics adopted in this study. Section 4 describes the proposed framework in detail, including hybrid dataset construction, detector comparison, model selection, retraining, and object-level bitemporal building comparison. Section 5 presents the experimental setup and data preparation procedures and discusses the comparative performance of the evaluated object detection models, the results of the selected final detector, and the findings of the candidate earthquake-related building-loss analysis. Finally, Section 6 summarizes the main findings, discusses the limitations of the proposed approach, and outlines directions for future research.

Table 1: Representative studies included in the literature synthesis

Ref.	Study / year	Imagery or dataset	Task	Method	Main outcome
[1]	Liu et al., 2022	Ground-level terrestrial images	Detection of cracks, spalling, debris, and collapse	Improved lightweight YOLOv5	Balanced accuracy, model size, and real-time inference for field deployment
[2]	Abdi et al., 2021	Post-event high-resolution imagery	Building damage recognition under limited labelled data	Deep transfer learning	Transfer learning improved damage recognition compared with training from scratch
[3]	Abdi et al., 2022	Post-event VHR orthophotos from multiple disaster types	Four-level damage classification: non-damage, minor, major, and collapsed	Fine-tuned pretrained ResNet-34	Demonstrated effective four-level classification using only post-event VHR imagery
[4]	Ji et al., 2018	Post-event QuickBird imagery, Haiti	Collapsed/non-collapsed building classification	SqueezeNet with over-sampling, under-sampling, and cost-sensitive learning	OA of 80.6% and 76.6% in two test areas; imbalance-aware strategies improved collapsed-building recognition
[5]	Ji et al., 2019	VHR satellite imagery, Haiti	Collapsed-building recognition	Comparison of handcrafted texture descriptors and CNN features	CNN features outperformed conventional texture-based representations
[6]	Ji et al., 2020	Pre/post VHR satellite imagery, Haiti	Earthquake-induced building destruction discrimination	Pretrained VGG-based CNN with transfer learning and augmentation	Transfer learning and augmentation improved destruction recognition
[7]	Ghaffarian et al., 2019	Multi-temporal WorldView-2 imagery and pre-disaster OSM building data	Post-disaster building database updating	ResUNet with Dense CRF and automatically generated OSM training masks	Automatically generated training samples and achieved an overall F1-score of approximately 84.2%
[8]	Zhan et al., 2022	Post-event aerial imagery, 2016 Kumamoto earthquake	Damaged-building instance extraction	Modified Mask R-CNN	Produced instance-level damaged-building maps from high-resolution imagery
[9]	Kalantar et al., 2020	Pre/post orthophotos	Earthquake-induced building damage classification	Comparison of CNN architectures and temporal-input strategies	Demonstrated the benefit of combining pre-event and post-event information

Continued on next page

369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399

Table 1 – continued from previous page

Ref.	Study / year	Imagery or dataset	Task	Method	Main outcome
[10]	Li et al., 2019	Post-event aerial imagery	Detection of undamaged buildings, damaged buildings, and ruins	SSD with VGG16-based convolutional-autoencoder pretraining and data augmentation	Demonstrated multi-class post-disaster object detection under limited labelled data
[11]	Ilmak et al., 2024	VHR satellite imagery, Antakya	Collapsed-building object detection	YOLOv7, YOLOv7x, YOLOv8l, and YOLOv8x	Reported YOLOv7 mAP@0.50 of approximately 0.79
[12]	Ali et al., 2020	Post-disaster satellite imagery	Weakly supervised destruction localisation	CNN trained with image-level destruction labels	Reduced dependence on pixel-level destruction masks
[13]	Tilon et al., 2020	xBD and UAV imagery	Unsupervised and transferable building-damage anomaly detection	Anomaly-detecting generative adversarial network trained on undamaged pre-disaster buildings	Reported best European test F1 of approximately 0.74
[14]	Duarte et al., 2019	Multi-temporal aerial oblique imagery	Building-façade damage classification	Mono-temporal and multi-temporal CNN fusion strategies	Multi-temporal fusion increased classification accuracy by up to 22%
[15]	Döş̄a et al., 2025	Pre/post VHR imagery and building footprints, Antakya	Footprint-supported building extraction and change analysis	SegNet, PSPNet, and ResUNet comparison	ResUNet achieved the strongest overall segmentation performance in the reported comparison
[16]	Xia et al., 2023	Ultra-high-resolution WorldView-2 imagery, İ̄slahiye	Building delineation and damage assessment for the 2023 Türkiye earthquake	BDANet with pre-event delineation and fused pre/post features	Demonstrated a specialised two-stage framework for earthquake building assessment
[17]	Xie et al., 2024	Google Earth pre/post imagery, Hatay	Object-level post-earthquake building assessment	YOLOv8 localisation, coordinate-based object pairing, HGC-assisted labelling, and R50M2Net classification	Developed object-level three-class assessment designed to tolerate view and illumination inconsistencies
[18]	Su et al., 2020	Bitemporal aerial RGB imagery, DSMs, and an old building map	Detection of new and demolished buildings	Deep multi-input change-detection network	Reported recall of 89.3% for new buildings and 88.8% for demolished buildings
[19]	Khodaverdi Zahraee & Rastiveis, 2017	Pre/post WorldView-2 imagery, Haiti	Object-level changed/unchanged building classification	Object segmentation, texture differences, and artificial neural network	Reported overall accuracy of 93%

Continued on next page

Table 1 – continued from previous page

Ref.	Study / year	Imagery or dataset	Task	Method	Main outcome
[20]	Tong et al., 2012	Pre/post high-resolution satellite stereo imagery, Wenchuan	Building-damage detection from geometric height change	Stereo-derived elevation and building-height difference analysis	Demonstrated collapsed-building detection using pre/post three-dimensional information
[21]	Gupta et al., 2019	xBD multi-disaster pre/post satellite imagery	Dataset development for building localisation and ordinal damage assessment	Expert-annotated building polygons and damage categories	Established a large-scale benchmark for building localisation and damage classification
[22]	Janalipour & Mohamadzadeh, 2019	Post-event LiDAR point clouds, DSM, and DTM	Automatic building-damage map production	Texture extraction and K-Means clustering combined with nDSM and building-inclination analysis	Training-data-free framework integrating texture, height, and inclination to identify different damage patterns
[23]	Zha et al., 2003	Landsat TM multi-spectral imagery	Automatic urban built-up area extraction	Normalized Difference Built-up Index (NDBI)	Established a spectral index for separating built-up areas from other land-cover classes
[24]	Alisjahbana et al., 2024	Multi-disaster satellite imagery derived from xBD	Building segmentation followed by damage classification	DeepDamageNet two-step deep-learning framework	Demonstrated a unified segmentation-and-classification pipeline
[26]	Miura et al., 2020	Post-disaster aerial imagery	Three-class assessment of collapsed, non-collapsed, and blue-tarp-covered buildings	CNN image-chip classification	Demonstrated multi-class post-disaster building assessment
[27]	Braik & Koliou, 2024	Georeferenced satellite imagery, GIS building data, and xBD-based training	Large-scale building damage mapping after Hurricane Ike	GIS-based building-chip extraction and CNN classification with application-area fine-tuning	Reported F1 of 0.88 on test data and 0.86 after fine-tuning on the Galveston application dataset
[28]	Wang et al., 2022	xBD pre/post satellite imagery	Building localisation followed by multi-level damage classification	Two-stage deep-learning framework with imbalance-aware learning	Improved damage classification on extremely imbalanced classes; classification-stage performance should not be interpreted as end-to-end detection performance
[35]	Bhangale et al., 2019	VHR remote-sensing imagery	Rapid earthquake damage detection	Deep CNN workflow with high-performance computing	Demonstrated rapid analysis of detailed VHR imagery

Continued on next page

431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461

462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492

Table 1 – continued from previous page

Ref.	Study / year	Imagery or dataset	Task	Method	Main outcome
[36]	Mazzanti et al., 2022	Sentinel-1 and COSMO-SkyMed SAR; 2016 Norcia earthquake	Post-earthquake damage detection and mapping	Amplitude and coherence change analysis at multiple SAR resolutions	VHR SAR supported building-level assessment, whereas Sentinel-1 mainly identified severely damaged sectors
[37]	Havivi et al., 2022	Pre/post COSMO-SkyMed SAR and Sentinel-2; Sulawesi	Rural damage mapping and characterisation of built-up and vegetation areas	Multi-sensor radar-optical change analysis	Demonstrated complementary use of SAR and optical observations for regional damage characterisation
[38]	Alataş & Taşkın, 2019	VHR post-event imagery of Bam	Preliminary building damage classification	Attribute Profiles and Multi-Attribute Profiles with k-NN	Morphological contextual features outperformed Haralick texture features
	Present study	Satellite-II Version 1 [25] plus 150 manually verified Google Earth images; independent 2022/2024 imagery	Single-class building detection and candidate demolition identification	SSD, Faster R-CNN, RetinaNet, YOLOv10, YOLOv11, and YOLOv12 benchmark followed by object-level temporal matching	Controlled cross-family benchmark and transferable temporal building-inventory workflow

2 Preliminaries

Before presenting the proposed framework, it is necessary to introduce the object detection architectures and evaluation metrics employed throughout this study. Because the objective is to identify the most suitable detector for large-scale building localization in historical Google Earth imagery, both one-stage and two-stage detection paradigms are considered. This section briefly reviews the theoretical foundations of SSD, Faster R-CNN, RetinaNet, and the YOLOv10, YOLOv11, and YOLOv12 architectures, followed by the evaluation metrics used for performance comparison and the specific challenges associated with building detection in high-resolution satellite and Google Earth imagery.

2.1 Single Shot MultiBox Detector (SSD)

The Single Shot MultiBox Detector (SSD) is a one-stage object detection architecture that performs object classification and bounding-box regression within a single forward pass. Unlike two-stage detectors, SSD does not require a separate region-proposal mechanism. Instead, object categories and location offsets are predicted directly from feature maps generated at different spatial resolutions. This multi-scale prediction strategy enables the model to process objects of varying sizes while maintaining relatively low inference latency. SSD assigns a set of predefined anchor boxes with different scales and aspect ratios to each feature-map location. During training, these anchors are matched with ground-truth objects according to their spatial overlap. The overall SSD objective is defined in Eq. (1) [30] as a weighted combination of classification and localization losses:

$$L_{\text{SSD}} = \frac{1}{N} (L_{\text{conf}} + \lambda_{\text{loc}} L_{\text{loc}}), \quad (1)$$

where N denotes the number of positive anchors matched to ground-truth objects, L_{conf} is the classification loss, L_{loc} is the localization loss, and λ_{loc} controls the contribution of the localization component.

The localization term in Eq. (1) can be expressed as:

$$L_{\text{loc}} = \sum_{i \in \text{pos}} \sum_{m \in \{x, y, w, h\}} \text{SmoothL1}(\hat{b}_i^m - b_i^m), \quad (2)$$

where $\hat{b}_i^m$ and b_i^m represent the predicted and reference parameters of the i -th bounding box, respectively. As indicated in Eq. (2), the regression component penalizes deviations in the center coordinates, width, and height of matched boxes.

The joint optimization defined in Eqs. (1) and (2) allows SSD to perform efficient end-to-end detection. Nevertheless, its dependence on predefined anchors and limited high-resolution semantic information may reduce its effectiveness for very small, densely distributed, or partially occluded buildings. In this study, SSD is therefore employed as a computationally efficient baseline against which more recent architectures are compared.

2.2 Faster R-CNN

Faster R-CNN follows a two-stage detection strategy in which candidate regions are generated before final object classification and bounding-box refinement. In the first stage, a Region Proposal Network estimates objectness probabilities and proposes candidate regions from convolutional feature maps. In the second stage, these proposals are classified and their coordinates are refined by the detection head.

The multi-task objective of the Region Proposal Network is expressed in Eq. (3) [29]:

$$L_{\text{RPN}} = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda_{\text{reg}} \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*), \quad (3)$$

where p_i denotes the predicted objectness probability of the i -th anchor, p_i^* is the corresponding ground-truth object/background label, t_i represents the predicted bounding-box transformation, and t_i^* is the reference transformation. The normalization terms N_{cls} and N_{reg} scale the classification and regression losses, whereas λ_{reg} balances their relative contributions.

The multiplicative term p_i^* in Eq. (3) ensures that bounding-box regression is performed only for positive anchors. This formulation enables the proposal network to learn objectness and localization jointly, while the second-stage detector performs more detailed classification and coordinate refinement. Faster R-CNN generally provides strong localization performance in complex scenes, particularly when objects are densely distributed. However, the region-proposal and region-processing stages increase computational cost and inference time. Accordingly, Faster R-CNN is used in this study as the representative two-stage detector for analysing the trade-off between localization accuracy and computational efficiency.

2.3 RetinaNet

RetinaNet is a one-stage detector designed to narrow the performance gap between dense one-stage detectors and two-stage architectures. The model combines a Feature Pyramid Network with focal loss to improve multi-scale representation and reduce the adverse effect of foreground-background imbalance.

The focal loss employed by RetinaNet is given in Eq. (4) [31]:

$$\text{FL}(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t), \quad (4)$$

where

$$p_t = \begin{cases} p, & y = 1, \\ 1 - p, & y = 0. \end{cases} \quad (5)$$

In Eqs. (4) and (5), p is the predicted probability of the positive class, α_t is a class-balancing factor, and γ is the focusing parameter. The modulating factor $(1 - p_t)^\gamma$ in Eq. (4) reduces the contribution of easily classified samples while preserving the influence of difficult or misclassified examples. This mechanism is particularly relevant in dense building detection, where background regions substantially outnumber positive building instances.

The multi-scale feature fusion performed by the Feature Pyramid Network can be represented in simplified form by Eq. (6):

$$P_l = \text{Conv}_{3 \times 3} [C_l + U_p(P_{l+1})], \quad (6)$$

where C_l denotes the backbone feature map at level l , P_l is the corresponding pyramid representation, and $U_p(\cdot)$ indicates upsampling from the next coarser level.

As expressed in Eq. (6), RetinaNet combines high-level semantic information with lower-level spatial detail. This design can improve the representation of small or difficult objects compared with SSD, while avoiding the explicit proposal stage required by Faster R-CNN. RetinaNet therefore occupies an intermediate position between computationally efficient one-stage detectors and highly accurate two-stage models.

2.4 YOLO Versions

The YOLO family has evolved through successive architectural developments aimed at improving object-detection accuracy, computational efficiency, and real-time inference capability. Although YOLOv10, YOLO11, and YOLOv12 all belong to the one-stage detection paradigm, they represent different design philosophies rather than merely consecutive software versions. Each model introduces distinct architectural strategies for feature extraction, prediction, and computational optimisation. The principal characteristics of the three architectures are summarized in Table 2 [32–34].

Table 2: Architectural comparison of YOLOv10, YOLO11, and YOLOv12

Feature	YOLOv10	YOLO11	YOLOv12
Detector paradigm	Real-time one-stage detector	Real-time one-stage detector	Real-time one-stage detector
Primary architectural emphasis	NMS-free end-to-end detection and efficiency-accuracy optimisation	Improved convolutional feature extraction and attention-assisted feature processing	Attention-centric real-time object detection
Main distinctive mechanism	Consistent Dual Assignments	C3k2, SPPF, and C2PSA modules	Area Attention and R-ELAN
Conventional NMS in the reported inference design	Not required for the end-to-end inference branch	Used in the standard implementation	Used in the reported implementation
Main stated contribution	Reducing redundant computation while enabling end-to-end prediction	Improving feature representation and the parameter-accuracy trade-off	Integrating attention mechanisms while retaining competitive real-time speed

As summarized in Table 2, YOLOv10 primarily focuses on improving computational efficiency by introducing Consistent Dual Assignments, which combine one-to-many supervision with one-to-one assignment during training. This strategy enables the dedicated inference branch to generate final predictions without relying on conventional Non-Maximum Suppression (NMS). Consequently, redundant post-processing operations are reduced while maintaining competitive detection accuracy [32].

YOLO11 follows a different architectural direction. Instead of redesigning the prediction pipeline, it strengthens feature extraction by incorporating C3k2, SPPF, and C2PSA modules. These components enhance multi-scale feature representation and attention-assisted feature learning while maintaining an appropriate balance between model complexity and detection performance. Therefore, the primary contribution

of YOLO11 lies in improving feature quality rather than modifying the inference mechanism [33].

YOLOv12 further extends this evolution by placing attention mechanisms at the centre of the detector architecture. As indicated in Table 2, the model introduces Area Attention together with Residual Efficient Layer Aggregation Networks (R-ELAN) to improve contextual feature interaction while preserving real-time inference capability. This design aims to strengthen feature representation in complex scenes without substantially increasing computational cost [34].

Although the three models employ different architectural components and optimisation strategies, they all perform object localisation and classification within a unified one-stage detection framework. Accordingly, the overall optimisation objective can be represented in a generalized implementation-level form as:

$$L_{\text{YOLO}} = \lambda_{\text{box}} L_{\text{box}} + \lambda_{\text{cls}} L_{\text{cls}} + \lambda_{\text{dff}} L_{\text{dff}}, \quad (7)$$

where L_{box} denotes the bounding-box regression loss, L_{cls} denotes the classification loss, and L_{dff} corresponds to the Distribution Focal Loss used for localization refinement. The weighting coefficients λ_{box} , λ_{cls} , and λ_{dff} determine the contribution of each loss component during optimization. Equation (7) is presented as a generalized implementation-level representation and should not be interpreted as the identical original loss function of YOLOv10, YOLO11, and YOLOv12, since their assignment strategies and architectural implementations differ [32–34].

Since only a single foreground category (building) is considered in this study, the classification component mainly separates buildings from the background. Consequently, detection performance depends primarily on feature representation, localisation accuracy, and the capability of identifying buildings with varying sizes, densities, and surrounding environments.

The spatial correspondence between a predicted bounding box B_p and its corresponding ground-truth box B_g is quantified using the Intersection over Union (IoU):

$$\text{IoU}(B_p, B_g) = \frac{|B_p \cap B_g|}{|B_p \cup B_g|}, \quad (8)$$

where higher IoU values indicate stronger agreement between predicted and reference bounding boxes. In the present study, IoU is employed both for matching predicted detections with reference objects and for evaluating localisation performance.

2.5 Evaluation Metrics

The performance of the evaluated object detection models was assessed using Precision, Recall, and F1-score. Precision quantifies the reliability of the predicted building detections by measuring the proportion of correct detections among all objects predicted as buildings. As defined in Eq. (9), a higher precision value indicates that the detector produces fewer false-positive building predictions.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (9)$$

Recall measures the detector’s ability to recover the buildings available in the reference annotations. According to Eq. (10), recall corresponds to the proportion of correctly detected buildings among all ground-truth building instances. A high recall value therefore indicates that relatively few buildings are missed by the model.

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (10)$$

Because precision and recall reflect different aspects of detection performance, the F1-score was used to summarize their balance within a single measure. As expressed in Eq. (11), the F1-score is calculated as the harmonic mean of precision and recall and decreases when either false-positive or false-negative detections become dominant.

$$F1\text{-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (11)$$

In Eqs. (9)–(11), TP denotes the number of correctly detected buildings, FP represents predicted bounding boxes that do not correspond to a reference building, and FN refers to ground-truth buildings that are not detected by the model. Considered together, the three measures provide complementary information about detection reliability, missed-building frequency, and the overall balance between these two error types.

2.6 Building Detection in High-Resolution Satellite and Google Earth Imagery

Building detection from high-resolution satellite and Google Earth imagery remains a challenging object detection problem due to the large variability in building appearance, scale, and surrounding environment. Unlike natural-image object detection, buildings observed from overhead imagery exhibit considerable differences in roof material, color, shape, orientation, and size. These variations become more pronounced in urban regions where buildings are densely distributed and frequently located in close proximity to one another.

Another challenge arises from the spatial resolution and acquisition characteristics of historical Google Earth imagery. Images acquired at different dates may present variations in illumination, viewing angle, seasonal conditions, atmospheric effects, and image quality. Consequently, the same building may exhibit substantial visual differences between acquisitions even when no structural change has occurred. Such inconsistencies complicate both building localization and subsequent temporal comparison.

Small buildings represent an additional difficulty. As object size decreases, the amount of discriminative visual information available to the detector becomes limited, while surrounding background regions occupy a relatively larger proportion of the receptive field. Roads, parking areas, bare soil, and concrete surfaces often exhibit spectral and textural characteristics similar to building roofs, increasing the likelihood

of false-positive detections. Furthermore, trees, shadows, neighboring structures, and partial occlusions may hide building boundaries and reduce localization accuracy.

Following major earthquakes, the detection task becomes even more challenging because collapsed buildings frequently lose their regular geometric appearance. Rubble, partially collapsed roofs, temporary shelters, construction activities, and debris fields may substantially alter the visual characteristics of affected areas. Consequently, post-event imagery often contains buildings that differ considerably from the building patterns learned during detector training.

For these reasons, the present study formulates building detection as a robust object detection problem rather than a conventional image classification task. The proposed framework evaluates multiple detector architectures under identical experimental conditions to identify the most appropriate model for reliable building localization before conducting temporal comparison.

3 Proposed Model

The overall workflow of the proposed method is illustrated in Fig. 1, whereas the data preparation and model-development procedures are formalized in Algorithms 1 and 2, respectively. The framework follows three consecutive stages: hybrid dataset construction, staged detector development, and bitemporal object-level comparison. Instead of applying direct pixel-wise differencing to historical Google Earth images, buildings are first localized independently at each acquisition date and are subsequently compared at the object level. This design separates localization errors from temporal matching errors and reduces the sensitivity of the analysis to illumination, seasonal variation, image compression, and minor geometric inconsistencies. The data preparation stage begins with the publicly available Satellite-II Version 1 building-detection dataset, which is supplemented with Google Earth imagery collected from areas affected by the 6 February 2023 Türkiye earthquakes. A total of 150 images acquired in 2022 and 150 corresponding images acquired in 2024 were collected from the same geographic locations, yielding 150 bitemporal image pairs. The additional Google Earth images were initially annotated with a previously trained single-class YOLO building detector. At this stage, the purpose of the model was only to generate provisional building bounding boxes; no image registration, homography estimation, damage classification, or temporal comparison was performed.

To improve sensitivity to small and densely distributed buildings, the automatic annotation procedure combined full-image inference with overlapping tiled inference. Each image was first processed as a whole. It was then divided into 1024×1024 pixel tiles with 25% overlap, and every tile was analysed independently. Tile-level detections were transformed back to the coordinate system of the original image and merged with the full-image predictions. Bounding boxes extending beyond image boundaries were clipped, extremely small detections were removed, and duplicate boxes generated by overlapping tiles were suppressed through a final non-maximum suppression operation. A deliberately low confidence threshold was used to prioritize recall during annotation, because false-positive boxes could subsequently be removed during manual inspection.

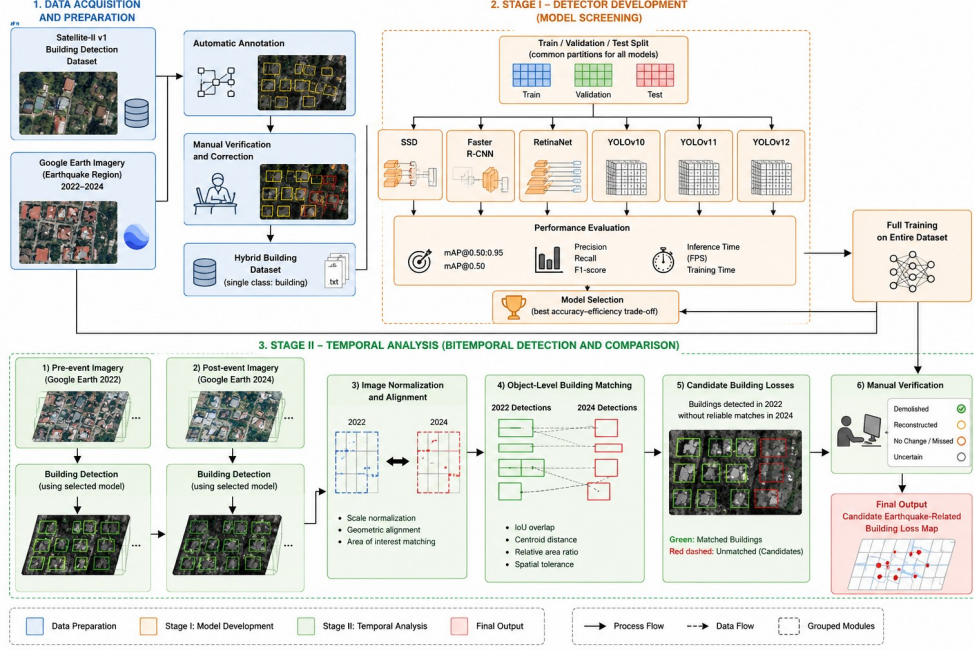


Fig. 1: Overall framework of the proposed method.

All valid detections were converted into the normalized single-class YOLO format, in which class 0 represented a building. A label file was generated for every image, including images for which no building was detected. Preview images showing the predicted boxes and a CSV-based annotation report were also produced to facilitate quality control. Automatic labels were not treated as final ground truth. Each image was manually reviewed; false-positive detections were removed, inaccurate bounding boxes were corrected, and missed buildings were added. The resulting annotations were therefore obtained through a semi-automatic workflow that combined the speed of model-assisted labelling with human verification. The complete annotation and partitioning procedure is summarized in Algorithm 1.

To prevent spatial data leakage, the dataset was partitioned at the paired-location level rather than at the individual-image level. Thus, the 2022 and 2024 images belonging to the same geographic location were always assigned to the same subset. Seventy location pairs, corresponding to 140 images, were allocated to training; 30 pairs, corresponding to 60 images, were assigned to validation; and the remaining 50 pairs, corresponding to 100 images, were reserved for testing. This grouping strategy prevents nearly identical geographic content from being shared across training and evaluation subsets.

Following dataset construction, SSD, Faster R-CNN, RetinaNet, YOLOv10, YOLOv11, and YOLOv12 were compared through a staged screening strategy. Because exhaustive full-dataset training of all candidate models would impose a substantial computational burden, the detectors were initially evaluated under a harmonized

```

815 Algorithm 1 Hybrid dataset construction and annotation
816 Require: Satellite-II Version 1 dataset  $D_S$ , Google Earth images  $G_{2022}$  and  $G_{2024}$ ,
817     pretrained single-class building detector  $M_{\text{pre}}$ 
818 Ensure: Training set  $D_{\text{train}}$ , validation set  $D_{\text{val}}$ , test set  $D_{\text{test}}$ 
819   1: Initialize paired-location dataset  $P \leftarrow \emptyset$ 
820   2: Load Satellite-II Version 1 dataset  $D_S$ 
821   3: Collect 150 Google Earth images from 2022 and the corresponding 150 images
822     from 2024
823   4: for each temporal image  $I \in \{G_{2022}, G_{2024}\}$  do
824     5:   Read image  $I$ 
825     6:   Perform full-image inference using  $M_{\text{pre}}$ 
826     7:   Divide  $I$  into overlapping tiles
827     8:   for each tile  $T$  in  $I$  do
828       9:     Detect candidate buildings in  $T$ 
829       10:    Transform tile boxes to original-image coordinates
830     11:   end for
831     12:   Merge full-image and tile-level detections
832     13:   Clip boxes and remove invalid or extremely small detections
833     14:   Apply final non-maximum suppression
834     15:   Convert detections into single-class YOLO labels
835     16:   if no building is detected then
836       17:     Save an empty label file
837     18:   end if
838     19:   Generate preview image and annotation report
839     20:   Manually remove false positives, correct boxes, and add missed buildings
840   21: end for
841   22: for each paired location  $i = 1, \dots, 150$  do
842     23:   Create  $P_i = \{I_{2022,i}, L_{2022,i}, I_{2024,i}, L_{2024,i}\}$ 
843     24:   Add  $P_i$  to  $P$ 
844   25: end for
845   26: Shuffle  $P$  using a fixed seed
846   27: Assign 70 pairs to training, 30 pairs to validation, and 50 pairs to testing
847   28: Verify that no location occurs in more than one subset
848   29: Expand pairs into 140 training, 60 validation, and 100 test images
849   30: return  $D_{\text{train}}, D_{\text{val}}, D_{\text{test}}$ 

```

851 protocol using the same training, validation, and test image identities and a limited training budget, while retaining architecture-specific input sizes and optimization requirements. Model performance was evaluated using mAP@0.50:0.95, mAP@0.50, precision, recall, and F1-score. Training time, average inference latency, and frames per second were additionally considered to characterize computational efficiency. The detector providing the most favorable accuracy-efficiency trade-off was selected and retrained using the complete hybrid dataset. Algorithm 2 summarizes the screening, selection, retraining, and temporal comparison stages.

Algorithm 2 Proposed framework for candidate earthquake-related building-loss detection	861
Require: Hybrid building dataset D and paired Google Earth images G_{2022} and G_{2024}	862
Ensure: Verified candidate earthquake-related building-loss map	863
1: Define $\mathcal{M} = \{\text{SSD, Faster R-CNN, RetinaNet, YOLOv10, YOLOv11, YOLOv12}\}$	864
2: for each detector $m \in \mathcal{M}$ do	865
3: Train m using the common training subset	866
4: Evaluate m on the validation and test subsets	867
5: Compute mAP@0.50:0.95, mAP@0.50, precision, recall, and F1-score	868
6: Measure training time, inference latency, and FPS	869
7: end for	870
8: Select m^* using the predefined accuracy-efficiency criterion	871
9: Retrain m^* using the complete hybrid dataset	872
10: for each paired location i do	873
11: Detect buildings $B_{2022,i}$ in the 2022 image	874
12: Detect buildings $B_{2024,i}$ in the 2024 image	875
13: Normalize image scale and estimate the optimal transformation using masked non-building structural references, phase correlation, and constrained ECC registration	876
14: Match buildings using IoU, buffered IoU, normalized centroid distance, and relative area consistency, followed by Hungarian one-to-one assignment	877
15: Store unmatched pre-event detections as candidate losses	878
16: end for	879
17: Manually review each candidate building loss	880
18: Remove candidates caused by missed detections, shadows, occlusion, reconstruction, or image differences	881
19: Generate the final verified candidate building-loss map	882
20: return Verified candidate building-loss map	883
	884
	885
	886
	887
	888
	889
	890
	891
The final detector was then applied independently to the 2022 and 2024 Google Earth images. To improve sensitivity to small and densely distributed buildings, detection combined full-image inference with overlapping tiled inference, followed by non-maximum suppression to merge duplicate detections. Before temporal comparison, corresponding image pairs were normalized to a common spatial scale. Detected building regions were masked during registration so that temporally more stable non-building structural features, including roads, intersections, parcel boundaries, and background edges, contributed more strongly to the alignment process. Candidate transformations based on size normalization, road/background phase correlation, and constrained ECC-based Euclidean and affine registration were evaluated, and the highest-scoring transformation was retained for each image pair. When sufficient surviving building correspondences were available, a small residual translation refinement based on building-centre agreement was optionally applied. Because the framework assigns the best available transformation to every same-area image pair, all 145 image pairs proceeded to temporal matching, resulting in 100% alignment coverage.	892
	893
	894
	895
	896
	897
	898
	899
	900
	901
	902
	903
	904
	905
	906

After alignment, the 2024 building detections were transformed into the coordinate system of the corresponding 2022 image. Object-level correspondence was then evaluated using complementary geometric evidence, including standard Intersection over Union (IoU), buffered IoU, normalized centroid distance, and relative bounding-box area consistency. These criteria were combined into a composite matching score, and global one-to-one correspondence between the 2022 and 2024 building inventories was established using the Hungarian algorithm. Buildings with reliable counterparts at both dates were treated as matched structures, whereas buildings detected in the 2022 imagery without a reliable 2024 counterpart were retained as candidate earthquake-related building losses.

An unmatched detection was not automatically interpreted as confirmed earthquake-induced demolition. Differences may also arise from controlled demolition, reconstruction, shadows, occlusion, vegetation, image-quality differences, residual geometric misalignment, or detector failure. Therefore, each automatically identified candidate was subsequently reviewed using the original bitemporal Google Earth imagery before inclusion in the final verified candidate building-loss inventory. This verification stage was used to reduce false temporal alarms while preserving the scalability of the detector-based workflow.

4 Simulation and Discussion

This section presents the experimental evaluation of the proposed framework. First, the construction of the hybrid dataset and the data preparation procedure are described. Subsequently, six representative object detection models are compared under identical experimental conditions to identify the most appropriate detector according to both detection accuracy and computational efficiency. After model selection, the selected detector is retrained using the complete hybrid dataset and evaluated on an independent Google Earth test set. Finally, the object-level bitemporal comparison results are presented to demonstrate the capability of the proposed framework for automatic screening of candidate earthquake-related building losses.

4.1 Experimental Dataset and Data Preparation

The experiments were conducted using a hybrid dataset constructed by combining the publicly available Satellite-II Version 1 dataset with a study-specific collection of historical Google Earth images. The Satellite-II dataset, obtained from Roboflow Universe, contains 2,400 satellite images annotated with single-class building bounding boxes and was used to provide the evaluated detectors with diverse examples of roof geometry, building scale, settlement density, and surrounding background conditions [25]. A representative Satellite-II image and its corresponding building annotations are shown in Fig. 2(a) and Fig. 2(b), respectively.

Because Satellite-II does not provide observations of the same geographic areas acquired at different dates, an additional Google Earth dataset was created to support the bitemporal component of the proposed framework. Images of the same 150 geographic locations were collected for both 2022 and 2024, resulting in 150 bitemporal image pairs and 300 Google Earth images in total. Representative images from

2024 and 2022 are presented in Fig. 2(c) and Fig. 2(e), respectively. These images were included not merely to increase the number of training samples, but to introduce the temporal information required for comparing building presence before and after the 6 February 2023 Türkiye earthquakes.

The Google Earth images were initially unlabeled. Preliminary building annotations were therefore generated using a previously trained single-class detector. To improve sensitivity to small and densely distributed buildings, full-image inference was combined with overlapping tiled inference. Tile-level detections were transformed back to the original image coordinates, merged with full-image predictions, and filtered using Non-Maximum Suppression to remove duplicate bounding boxes. The resulting detections were converted into the YOLO annotation format.

All automatically generated annotations were subsequently inspected manually. False-positive detections were removed, inaccurate bounding boxes were refined, and buildings missed during automatic labeling were added. Examples of the resulting manually verified annotations for the 2024 and 2022 images are shown in Fig. 2(d) and Fig. 2(f), respectively. The comparison between the original and annotated images illustrates the contribution of the semi-automatic labeling procedure to the construction of the study-specific dataset.

To prevent spatial data leakage, the Google Earth images were partitioned at the geographic-pair level rather than at the individual-image level. Thus, the 2022 and 2024 images representing the same location were always assigned to the same subset. Of the 150 bitemporal pairs, 70 pairs (140 images) were allocated to training, 30 pairs (60 images) to validation, and 50 pairs (100 images) to the independent test set. The Google Earth training subset was combined with the Satellite-II training data to form the hybrid training set, whereas the validation and test subsets were excluded from parameter optimization. The validation set was used to monitor convergence and select the optimal checkpoint, while the test set was reserved exclusively for final performance evaluation.



Fig. 2: Representative examples from the hybrid dataset and the corresponding building annotations.



(e) Google Earth 2022



(f) Annotation

Fig. 2: Representative examples from the hybrid dataset and the corresponding building annotations (continued): (a) an original Satellite-II image, (b) its building annotations, (c) an unannotated Google Earth image acquired in 2024, (d) the corresponding manually verified building annotations, (e) an unannotated Google Earth image acquired in 2022, and (f) the corresponding manually verified building annotations.

1091 4.2 Comparative Evaluation of Object Detection Models

1092 To ensure a fair and unbiased comparison, all object detection models were trained
1093 and evaluated under an identical experimental protocol. The comparison included
1094 SSD, Faster R-CNN, RetinaNet, YOLOv10, YOLOv11, and YOLOv12, each trained
1095 using the same hybrid dataset, identical training, validation, and test partitions, and
1096 the same random seed. All experiments were performed using an input image reso-
1097 lution of 320×320 pixels, a batch size of 4, a maximum of 30 training epochs, and
1098 an AdamW optimizer with identical optimization parameters. Early stopping with a
1099 patience value of five epochs was employed to avoid unnecessary training after conver-
1100 gence. To eliminate variations caused by dataset composition, all models used exactly
1101 the same training, validation, and testing images. Data augmentation settings, includ-
1102 ing geometric and photometric transformations, were also kept identical throughout
1103 the experiments. Consequently, the observed performance differences can be attributed
1104 primarily to the architectural characteristics of the evaluated detectors rather than to
1105 differences in data preparation or training configuration. Following training, all detec-
1106 tors were evaluated on the same independent test set using identical evaluation metrics
1107 and inference settings.

1108 Table 3 presents the building-detection performance of the six evaluated archi-
1109 tectures on the independent test set. The test set contained 3,978 reference building
1110 instances, as confirmed by the identical $TP+FN$ value obtained for all models. Among
1111 the evaluated detectors, YOLOv12n achieved the strongest overall detection perfor-
1112 mance, obtaining the highest recall of 39.29%, the highest F1-score of 44.97%, and the
1113 highest mAP@0.50 of 37.91%. It correctly detected 1,563 buildings while producing
1114 1,411 false-positive detections and missing 2,415 reference buildings. These findings
1115 indicate that YOLOv12n provided the most balanced performance between recovering
1116 existing buildings and limiting incorrect detections.

1117 YOLOv11n achieved the highest precision, reaching 59.39%. It produced the lowest
1118 number of false-positive detections among the YOLO variants, with 978 false positives.
1119 However, its recall was lower than those of YOLOv10n and YOLOv12n, indicating
1120 that its more conservative predictions resulted in a larger number of missed buildings.
1121 Nevertheless, its F1-score of 44.79% remained very close to that of YOLOv12n.

1122 YOLOv10n also demonstrated competitive performance, with a recall of 39.14%,
1123 an F1-score of 43.37%, and an mAP@0.50 of 36.67%. Although its recall was close to
1124 that of YOLOv12n, its higher false-positive count reduced its precision to 48.63%.

1125 Among the conventional architectures, Faster R-CNN achieved the best overall
1126 performance, with a precision of 49.32%, a recall of 30.27%, an F1-score of 37.51%,
1127 and an mAP@0.50 of 27.46%. However, its performance remained below that of all
1128 three YOLO variants.

1129 SSD detected 1,492 reference buildings and achieved a recall of 37.51%, which was
1130 relatively close to the YOLO models. Nevertheless, it generated 6,133 false-positive
1131 detections, resulting in a low precision of 19.57% and an F1-score of 25.72%. This
1132 result indicates that SSD detected a considerable proportion of buildings but produced
1133 many incorrect predictions.

1134 RetinaNet showed the weakest performance under the applied experimental condi-
1135 tions. It correctly detected only 195 buildings, missed 3,783 buildings, and generated
1136

8,105 false-positive detections. Consequently, its precision, recall, F1-score, and mAP@0.50 remained at 2.35%, 4.90%, 3.18%, and 0.16%, respectively.

Overall, the three YOLO models substantially outperformed Faster R-CNN, SSD, and RetinaNet. Since missed building detections may be incorrectly interpreted as candidate building losses during the subsequent bitemporal analysis, recall and balanced detection performance were considered particularly important. Therefore, YOLOv12n was selected for full-scale training and subsequent bitemporal building comparison.

Table 3: Detection performance of the evaluated object detection models on the independent test set

Model	TP	FP	FN	Precision (%)	Recall (%)	F1 Score (%)	mAP@0.50 (%)
Faster R-CNN	1204	1237	2774	49.32	30.27	37.51	27.46
RetinaNet	195	8105	3783	2.35	4.90	3.18	0.16
SSD	1492	6133	2486	19.57	37.51	25.72	12.51
YOLOv10n	1557	1645	2421	48.63	39.14	43.37	36.67
YOLOv11n	1430	978	2548	59.39	35.95	44.79	36.93
YOLOv12n	1563	1411	2415	52.56	39.29	44.97	37.91

Table 4 compares the computational efficiency and architectural complexity of the evaluated models. YOLOv10n was the fastest model, with a mean inference time of 48.80 ms per image and a throughput of 20.49 FPS. YOLOv11n closely followed with 50.86 ms per image and 19.66 FPS, whereas YOLOv12n achieved an inference time of 57.89 ms and 17.27 FPS. Although YOLOv12n was slightly slower than the other two YOLO variants, it remained substantially faster than the conventional detectors.

Faster R-CNN exhibited the highest inference latency, requiring 306.68 ms per image. RetinaNet and SSD required 230.88 ms and 254.73 ms per image, respectively. Accordingly, the YOLO-based models operated approximately four to six times faster than the conventional architectures.

The YOLO models also provided substantial advantages in terms of model size and parameter count. Their weight files were approximately 5 MB, with parameter counts of approximately 2.6–2.7 million. In contrast, Faster R-CNN had a model size of 158.05 MB and more than 41 million parameters. RetinaNet and SSD were also considerably larger and computationally more demanding than the YOLO variants.

With respect to training time, YOLOv11n completed training in the shortest duration, followed by YOLOv10n and YOLOv12n. Although YOLOv12n required slightly more training time than the other YOLO variants, it provided the highest overall detection performance while retaining a small model size and acceptable inference speed. It was therefore considered the most appropriate model for accuracy-oriented full-scale retraining and subsequent bitemporal building analysis.

As shown in Fig. 3(a), the YOLO-based detectors achieved substantially better overall detection performance than Faster R-CNN, SSD300, and RetinaNet. YOLOv11n produced the highest precision of 59.39%, indicating the strongest ability to suppress false-positive detections. However, YOLOv12n achieved the highest recall

Table 4: Computational efficiency, training time, and model complexity of the evaluated detectors

Model	Mean inference time (ms/image)	Inference time SD (ms)	FPS	Training time (s)	Model size (MB)	Parameters
Faster R-CNN	306.68	16.14	3.26	293205.80	158.05	41352281
RetinaNet	230.88	10.096	4.33	246436.00	123.24	32221814
SSD	254.73	14.31	3.93	157206.90	90.61	23745908
YOLOv10n	48.80	2.51	20.49	54708.94	5.44	2707430
YOLOv11n	50.86	5.95	19.66	31630.43	5.17	2590035
YOLOv12n	57.89	5.10	17.27	55104.99	5.21	2568243

of 39.29% and the highest F1-score of 44.97%, demonstrating the best balance between detecting reference buildings and limiting incorrect predictions.

The localization results in Fig. 3(b) further support this finding. YOLOv12n obtained the highest mAP@0.50 value of 37.91%, whereas YOLOv10n achieved the highest mAP@0.50:0.95 value of 15.72%. The difference between YOLOv10n and YOLOv12n at the stricter mAP@0.50:0.95 criterion was only 0.30 percentage points, indicating that YOLOv12n retained competitive localization performance over multiple IoU thresholds.

The heatmap in Fig. 3(c) provides an integrated representation of these results and clearly highlights the superior performance of the YOLO family, particularly YOLOv12n, across the principal detection metrics. The TP, FP, and FN distributions in Fig. 3(d) provide further insight into model behavior. YOLOv12n produced the highest number of true-positive detections, correctly identifying 1,563 buildings, while YOLOv11n generated the lowest number of false positives among the YOLO variants. In contrast, SSD300 and RetinaNet produced considerably larger numbers of false-positive detections, which substantially reduced their precision and overall F1 performance.

Collectively, the accuracy-based results indicate that YOLOv12n achieved the best overall building-detection performance and was therefore selected for full-scale training and subsequent bitemporal analysis.

Figure 4 summarizes the computational performance of the evaluated detectors. As shown in Fig. 4(a), YOLOv10n achieved the lowest mean inference time, requiring 48.80 ms per image, followed by YOLOv11n with 50.86 ms and YOLOv12n with 57.89 ms. In comparison, Faster R-CNN, SSD300, and RetinaNet required substantially longer inference times ranging from approximately 231 to 307 ms per image. Thus, all three YOLO variants were approximately four to six times faster than the conventional detectors.

The relationship between accuracy and speed is illustrated in Fig. 4(b), where the YOLO models are clustered in the most favorable region of the plot. YOLOv10n provided the highest computational efficiency and the highest mAP@0.50:0.95, whereas YOLOv12n combined slightly lower speed with the strongest overall detection accuracy. Although YOLOv12n was slower than YOLOv10n and YOLOv11n, its inference time remained suitable for large-scale image analysis and was substantially lower than

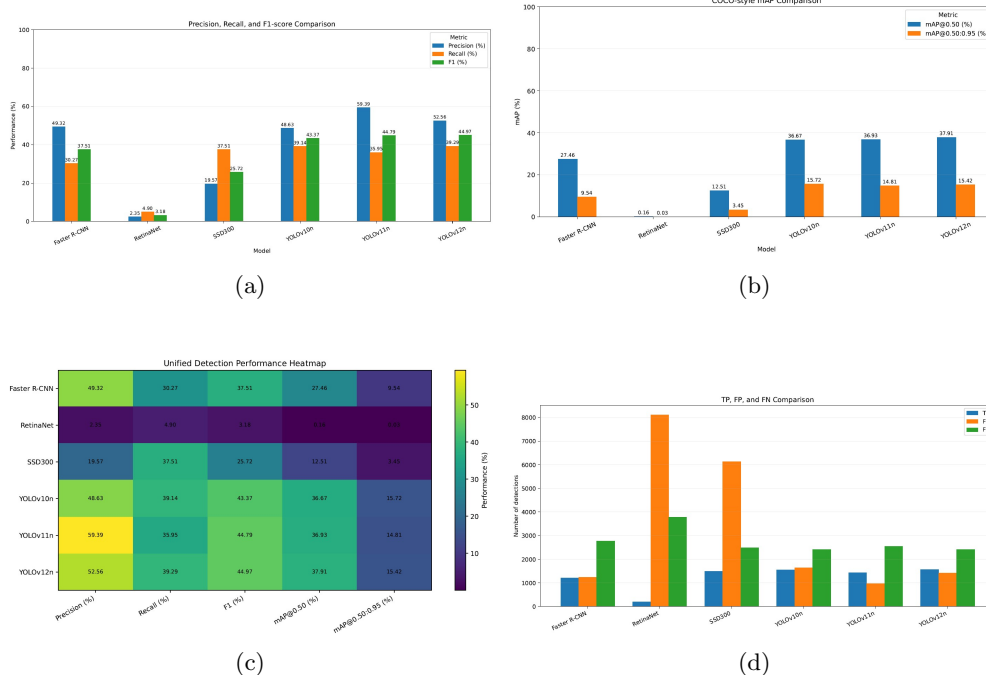


Fig. 3: Comparative detection performance of the evaluated object detection models: (a) precision, recall, and F1-score; (b) mAP@0.50 and mAP@0.50:0.95; (c) overall performance heatmap; and (d) true positive, false positive, and false negative detection counts.

those of the conventional architectures. Therefore, the modest computational cost of YOLOv12n was considered acceptable given its superior recall, F1-score, mAP@0.50, and true-positive detection count.

4.3 Final Model Selection and Detection Results

Section 5.2 presented a comparative evaluation of six object detection models under identical experimental conditions. To ensure a fair comparison while keeping the computational burden manageable, the comparative experiments were conducted using 50% of the hybrid training dataset and a fixed training schedule of 30 epochs. Even under this reduced protocol, the computational cost of some conventional architectures remained substantial. Faster R-CNN required approximately 293,205.8 s, RetinaNet 246,436.0 s, and SSD 157,206.9 s for training, whereas YOLOv12n completed training in 55,104.99 s. In addition, YOLOv12n contained only 2.57 million parameters, compared with approximately 41.35 million for Faster R-CNN, 32.22 million for RetinaNet, and 23.75 million for SSD. These results indicated that YOLOv12n provided a markedly more favorable computational profile than the conventional detectors while maintaining competitive detection performance.

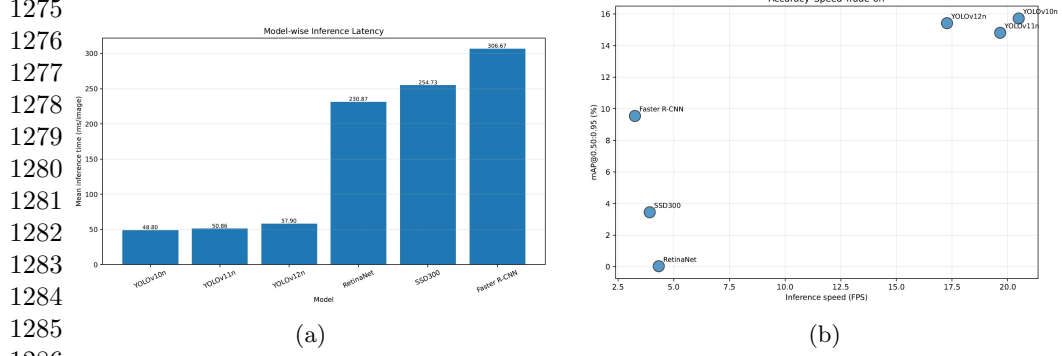


Fig. 4: Computational performance of the evaluated object detection models: (a) mean inference time per image and (b) accuracy–speed trade-off based on FPS and mAP@0.50:0.95.

Considering that even the reduced training protocol resulted in long training times for several architectures, repeating full-scale training using the complete hybrid dataset for all six models was considered computationally impractical. Therefore, the experiments in Section 5.2 were used primarily for architecture selection, and only the best-performing detector, YOLOv12n, was subsequently subjected to full training. This staged strategy enabled a controlled comparison of detector architectures while reserving the computationally more demanding full-data training procedure for the model offering the most favorable accuracy–efficiency trade-off.

Following model selection, YOLOv12n was retrained using the complete hybrid training dataset described previously. In contrast to the comparative stage, which employed only half of the available training data and a fixed 30-epoch schedule, the final training used the entire training subset and an extended schedule of up to 100 epochs. The training configuration also incorporated early stopping with a patience value of 15 epochs and a broader augmentation strategy including rotation, translation, scaling, shearing, perspective transformation, horizontal and vertical flipping, and HSV-based photometric augmentation. The best weights obtained in the preceding stage were used as the initialization point so that the detector could further adapt to the visual characteristics of the manually corrected 2022 and 2024 Google Earth imagery.

The final detector was then evaluated on an independent Google Earth test subset consisting of 100 previously unseen images corresponding to 50 geographic image pairs acquired in 2022 and 2024. These images were excluded from both model fitting and checkpoint selection, thereby providing an independent assessment of the final detector. The corresponding confusion matrices are presented in Fig. 5, while the comparative and full-training performance metrics are summarized in Table 5.

As shown in Fig. 5(a), the final YOLOv12n detector correctly identified 584 building instances, while 1,172 reference buildings were not detected and 225 background regions were incorrectly predicted as buildings. The normalized confusion matrix in Fig. 5(b) further indicates that the dominant error type was associated with missed

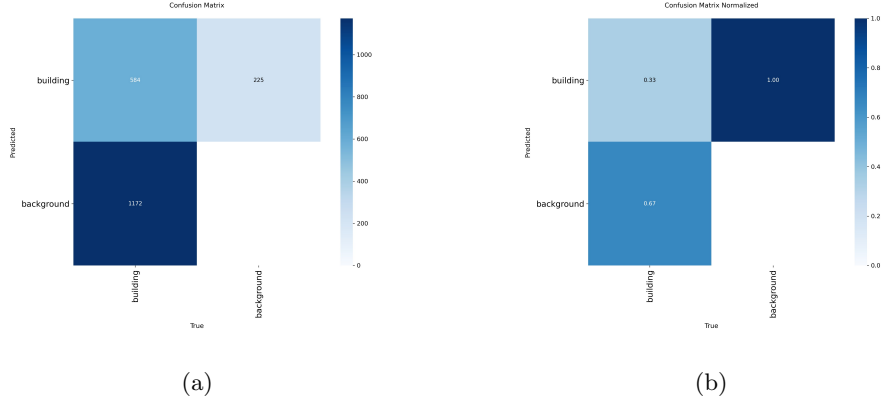


Fig. 5: Confusion-matrix results of the final YOLOv12n model on the independent Google Earth test set: (a) confusion matrix based on absolute detection counts and (b) normalized confusion matrix.

Table 5: Comparison of YOLOv12n detection performance between the comparative training and full-training stages on the independent Google Earth test set

Model	Precision (%)	Recall (%)	F1 Score (%)
YOLOv12n – comparative training	52.56	39.29	44.97
YOLOv12n – full training	72.19	33.26	45.54

building instances rather than false-positive predictions. Approximately 33% of the reference building instances were correctly detected, whereas the remaining proportion was assigned to the background category.

The performance values reported in Table 5 provide additional insight into the effect of full-data training. Compared with the reduced comparative-training stage, the final YOLOv12n model exhibited a substantial increase in precision, from 52.56% to 72.19%, while the F1-score increased slightly from 44.97% to 45.54%. In contrast, recall decreased from 39.29% to 33.26%. This pattern indicates that the final model became more conservative after full training: false-positive predictions were reduced considerably, but this improvement in precision was accompanied by a higher proportion of missed building instances.

Accordingly, the full-training stage did not improve all detection metrics simultaneously; rather, it shifted the detector toward more reliable positive predictions while maintaining a similar overall balance between precision and recall. This behavior should be interpreted in the context of the substantial visual complexity of the Google Earth imagery used in this study. Compared with conventional benchmark datasets, historical Google Earth images acquired over earthquake-affected urban areas exhibit

considerable variation in viewing geometry, spatial resolution, illumination, seasonal conditions, shadows, compression artifacts, and local image quality.

In addition, post-earthquake buildings may partially or completely lose their regular roof geometry because of collapse, debris accumulation, reconstruction activities, or occlusion. Small and densely distributed buildings are particularly difficult to localize because their boundaries may be weak or partially indistinguishable from adjacent roads, bare ground, rubble, and neighboring structures.

Despite these challenges, the full-data training stage substantially improved the precision of YOLOv12n and enabled the detector to adapt more directly to the study-specific Google Earth imagery. The resulting model was therefore used to generate the final building inventories for the subsequent bitemporal analysis. Because the reduced recall indicates that some buildings may still remain undetected, the temporal-analysis stage was not based solely on the absence of post-event detections. Instead, spatial alignment, object-level geometric matching, and subsequent verification of unmatched candidates were incorporated to reduce the risk of interpreting missed detections as actual building losses.

Figure 6 illustrates the variation in the performance of the final YOLOv12n model under different confidence thresholds. As shown in Fig. 6(a), the F1-score reached its maximum at a confidence threshold of approximately 0.14 and gradually decreased as the confidence threshold increased, primarily due to the reduction in recall. Figure 6(b) demonstrates that precision increased with increasing confidence threshold, approaching 1.0 at higher confidence values. In contrast, Fig. 6(d) shows a continuous decline in recall as the confidence threshold increased, reflecting the expected trade-off between precision and recall.

The overall discriminative performance of the model is summarized by the Precision–Recall curve presented in Fig. 6(c). The area under the curve ($\text{mAP}@0.50 \approx 0.45$) indicates that the model achieved a satisfactory level of building detection performance on the Google Earth imagery acquired over the earthquake-affected regions. When considered together with the Precision–Recall relationship, the results indicate that the model adopted a more conservative detection strategy at higher confidence thresholds, substantially reducing false-positive detections while increasing the likelihood of missed building instances.

These observations are consistent with the confusion-matrix results presented in Fig. 5 and the quantitative performance metrics summarized in Table 5. Overall, the final YOLOv12n model demonstrated more reliable positive predictions following full-data training; however, detecting small, densely distributed, partially occluded, or severely damaged buildings remained a challenging task due to the high visual complexity and temporal variability of the Google Earth imagery.

4.4 Bitemporal Analysis

Following the final building detection stage, the corresponding Google Earth images acquired in 2022 and 2024 were analyzed using the proposed object-level bitemporal matching framework. The study focused on selected districts in Hatay and Kahramanmaraş, which were among the regions most severely affected by the 6 February 2023 Türkiye earthquakes. The investigated districts included Antakya, Defne, Hassa,

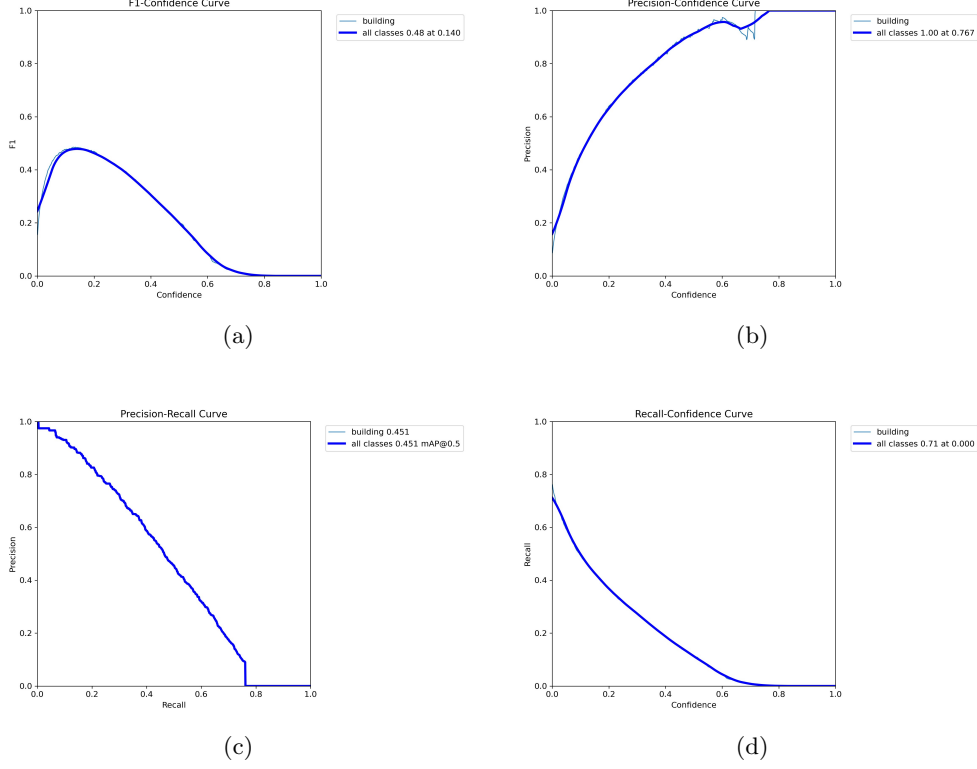


Fig. 6: Performance curves of the final YOLOv12n model evaluated on the independent Google Earth test set: (a) F1–Confidence curve, (b) Precision–Confidence curve, (c) Precision–Recall curve, and (d) Recall–Confidence curve.

Kırıkhan, Samandağ, İskenderun, Kahramanmaraş (Central), Onikişubat, and Dulka-diroğlu. These study areas were selected because they experienced extensive structural damage and represent diverse urban environments with varying building densities, settlement patterns, urban morphology, and post-earthquake physical changes. Furthermore, historical Google Earth imagery acquired in both 2022 (pre-earthquake) and 2024 (post-earthquake) was available for the same geographic locations, enabling reliable object-level temporal comparison.

Before object-level matching, each 2024 image was geometrically aligned to its corresponding 2022 image. To minimize the influence of actual building destruction on image registration, detected building regions were excluded during the registration process, allowing temporally stable spatial features, such as roads, road intersections, parcel boundaries, and background edges, to contribute more effectively to image alignment. The registration framework combined size normalization, road/background-based phase correlation, ECC-based Euclidean and affine transformations, and residual refinement based on surviving building centers whenever appropriate.

A total of 145 bitemporal image pairs were processed, and an alignment solution was successfully obtained for all image pairs. The detector identified 2,872 buildings in the 2022 imagery and 2,039 buildings in the corresponding 2024 imagery. Following spatial registration and one-to-one object matching, 842 buildings detected in 2022 had reliable counterparts in the 2024 imagery, whereas 2,030 buildings remained unmatched and were automatically labelled as candidate building losses.

Table 6: Overall results of the automatic bitemporal candidate building-loss screening

Metric	Value
Image pairs	145
Successful registrations	145
Registration success rate (%)	100.00
Buildings detected in 2022	2,872
Buildings detected in 2024	2,039
Matched buildings	842
Candidate building losses	2,030
Candidate loss rate (%)	70.68

As summarized in Table 6, 29.32% of the buildings detected in the 2022 imagery were successfully matched with corresponding buildings in the 2024 imagery, while the remaining 70.68% were retained as initial automatic candidate building losses. This percentage should not be interpreted as the confirmed earthquake-induced building-loss rate. Instead, it represents the output of the automatic screening stage, in which unmatched buildings are identified as locations requiring further investigation. The overall composition of the pre-earthquake building inventory following temporal matching is illustrated in Fig. 7.

As shown in Fig. 7, of the 2,872 buildings detected in the 2022 imagery, 842 buildings (29.32%) were successfully matched in the corresponding 2024 imagery, whereas 2,030 buildings (70.68%) remained within the automatic candidate building-loss category. It should be emphasized that these unmatched buildings do not necessarily represent confirmed earthquake-induced building losses. Instead, they constitute the initial output of the automatic screening framework and require subsequent manual verification. The spatial distribution of automatically identified candidate building losses across the investigated districts is presented in Fig. 8, whereas the corresponding numerical results are summarized in Table 7.

According to Fig. 8(a), Antakya exhibited the largest number of automatically identified candidate building losses (715 buildings), followed by Kahramanmaraş (Central) (513), Defne (224), and Onikişubat (223). Lower candidate counts were observed in Kırıkhan (170), Dulkadiroğlu (74), İskenderun (73), Hassa (37), and Samandağ (1).

Composition of the 2022 building inventory after temporal matching

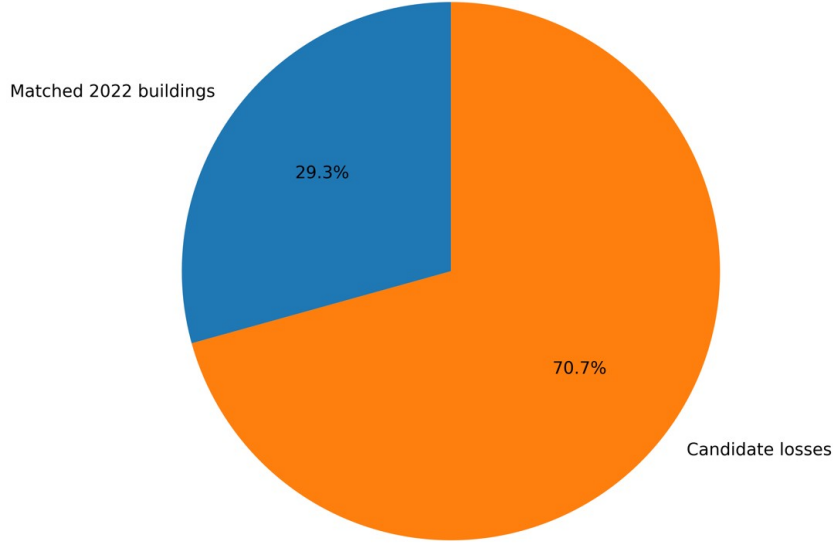


Fig. 7: Composition of the 2022 detected-building inventory after object-level temporal matching.

When normalized with respect to the pre-earthquake building inventory, Fig. 8(b) shows that Antakya also exhibited the highest initial automatic candidate building-loss rate (86.14%), followed by Kahramanmaraş (Central) (78.68%) and Defne (77.24%). Candidate-loss rates of 57.18%, 52.86%, 52.11%, 50.00%, 49.71%, and 47.10% were observed for Onikişubat, Dulkadiroğlu, Hassa, Samandağ, Kırıkhan, and İskenderun, respectively.

Table 7: District-level building inventories, temporal matching results, and initial automatic candidate building-loss rates

District	Image pairs	Buildings 2022	Buildings 2024	Matched buildings	Candidate losses	Candidate loss rate (%)	Matched rate (%)
Antakya	47	830	390	115	715	86.14	13.86
Kahramanmaraş (Central)	8	652	195	139	513	78.68	21.32
Defne	29	290	240	66	224	77.24	22.76
Onikişubat	22	390	334	167	223	57.18	42.82
Kırıkhan	15	342	429	172	170	49.71	50.29
Dulkadiroğlu	6	140	132	66	74	52.86	47.14
İskenderun	10	155	187	82	73	47.10	52.90
Hassa	5	71	79	34	37	52.11	47.89
Samandağ	3	2	53	1	1	50.00	50.00

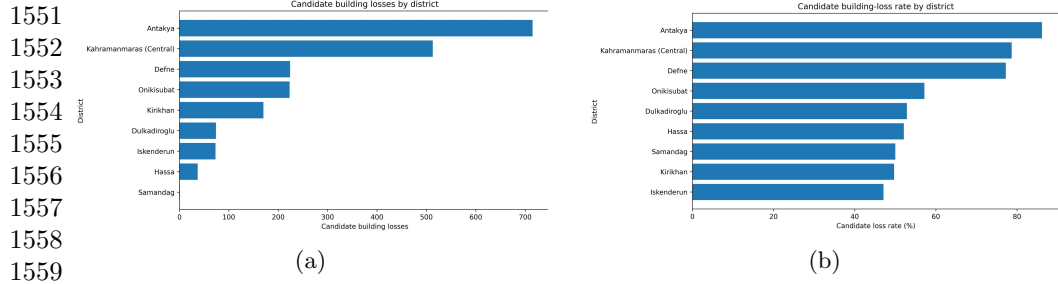


Fig. 8: District-level results of the automatic candidate building-loss screening: (a) number of automatically identified candidate building losses and (b) initial automatic candidate building-loss rate.

As shown in Table 7, the high candidate-loss values observed in Antakya, Kahramanmaraş (Central), and Defne were evident not only in absolute counts but also in the normalized rates. However, district-level percentages should be interpreted together with the size of the underlying building inventory. For example, the 50% candidate-loss rate observed in Samandağ was calculated from only two buildings detected in 2022 and therefore should not be interpreted with the same evidential weight as rates derived from several hundred buildings.

The confidence-score distribution of automatically identified candidate building losses is presented in Fig. 9, while district-level descriptive statistics are summarized in Table 8.

The distribution shown in Fig. 9 indicates that candidate detections were concentrated predominantly in the lower-to-moderate confidence range, although the distribution extended to substantially higher confidence values. This finding suggests that unmatched detections were not limited exclusively to low-confidence predictions near the detection threshold.

Table 8: District-level confidence statistics of automatically identified candidate building losses

District	Candidate count	Mean confidence	SD confidence	Median confidence	Min confidence	Max confidence
Antakya	715	0.319	0.094	0.299	0.200	0.721
Kahramanmaraş (Central)	513	0.369	0.130	0.336	0.200	0.734
Defne	224	0.306	0.084	0.288	0.200	0.587
Onikişubat	223	0.312	0.087	0.294	0.200	0.561
Kırıkhan	170	0.319	0.107	0.292	0.200	0.734
Dulkadiroğlu	74	0.305	0.083	0.291	0.200	0.604
İskenderun	73	0.286	0.081	0.259	0.200	0.556
Hassa	37	0.284	0.078	0.256	0.202	0.499
Samandağ	1	0.339	—	0.339	0.339	0.339

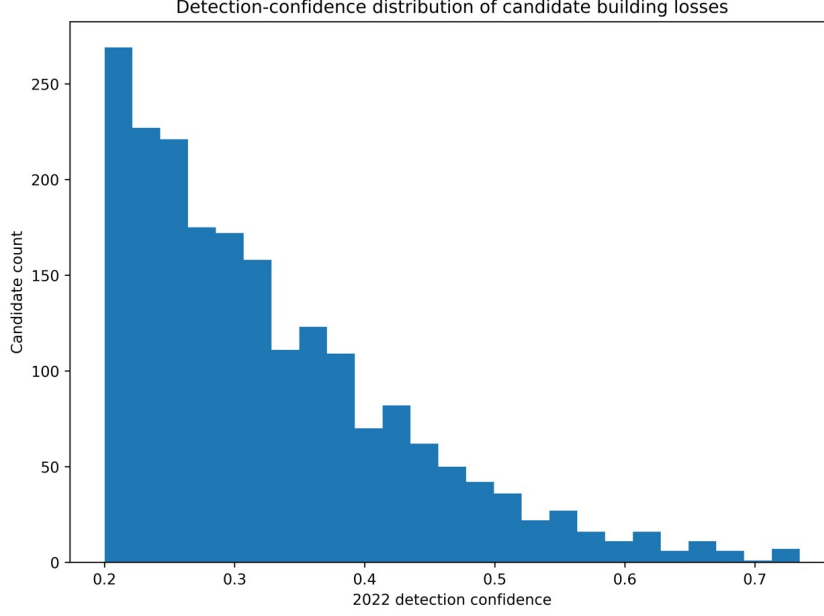


Fig. 9: Confidence-score distribution of automatically identified candidate building losses.

As summarized in Table 8, mean confidence values ranged from 0.284 to 0.369 across districts. The highest mean confidence was observed in Kahramanmaraş (Central) (0.369), whereas the lowest values were recorded in Hassa (0.284) and İskenderun (0.286). Antakya, which contained the largest candidate set, had a mean confidence of 0.319 ± 0.094 and a median value of 0.299.

These results indicate that candidate building losses were not restricted solely to extremely low-confidence detections; nevertheless, confidence scores alone are insufficient to distinguish actual building disappearance from detection or registration errors. Object-level correspondence was established by combining standard IoU, buffered IoU, normalized centroid distance, and relative bounding-box area consistency, followed by global one-to-one assignment using the Hungarian algorithm.

Importantly, unmatched 2022 detections were explicitly stored as candidate losses requiring manual review, rather than being automatically classified as earthquake-induced demolitions. Accordingly, the 2,030 automatically identified candidate buildings and the 70.68% initial automatic candidate-loss rate should be interpreted as the outcome of a high-throughput screening stage. Unmatched detections may reflect actual building collapse, planned demolition, reconstruction, footprint modification, missed post-event detection, occlusion, vegetation, shadows, image-quality differences, or residual registration errors. Therefore, the proposed framework is intended to reduce the number of locations requiring detailed expert inspection rather than replace manual verification. Following visual assessment of the original 2022 and 2024 imagery, the

retained cases can subsequently be reported separately as verified candidate building losses.

5 Conclusions

This study proposed an object-level bitemporal framework for automatically identifying candidate building losses from historical Google Earth imagery acquired before and after the 6 February 2023 Türkiye earthquakes. Unlike many previous studies that depend on building footprints, GIS databases, LiDAR data, digital surface models, semantic segmentation, or manually prepared building inventories, the proposed framework relies solely on historical Google Earth imagery combined with deep-learning-based building detection, robust image registration, and object-level temporal matching. The principal findings of this study are summarized as follows:

1. A comprehensive comparison of six state-of-the-art object detectors (Faster R-CNN, RetinaNet, SSD, YOLOv10n, YOLOv11n, and YOLOv12n) demonstrated that YOLOv12n achieved the most favorable trade-off between detection accuracy and computational efficiency. Under identical experimental conditions, YOLOv12n achieved an F1-score of 44.97%, Recall of 39.29%, and mAP@0.50 of 37.91%, while requiring only 2.57 million parameters and a model size of 5.21 MB.
2. The selected YOLOv12n detector was retrained using the complete hybrid dataset, which combined the publicly available Satellite-II dataset with manually verified Google Earth imagery. Evaluation on an independent test set consisting of 100 previously unseen Google Earth images yielded a Precision of 72.19%, Recall of 33.26%, and F1-score of 45.54%. Compared with the initial comparative training, the final model substantially reduced false-positive detections while maintaining comparable overall detection performance.
3. The proposed bitemporal framework successfully processed all 145 image pairs, corresponding to a 100% image-registration success rate. The detector identified 2,872 buildings in the 2022 imagery and 2,039 buildings in the corresponding 2024 imagery.
4. Object-level temporal matching successfully associated 842 buildings across the two acquisition dates, whereas 2,030 buildings remained unmatched and were automatically identified as candidate building losses, corresponding to an initial automatic candidate building-loss rate of 70.68% relative to the detected 2022 building inventory.
5. Significant spatial variation was observed among districts. Antakya exhibited the largest number of candidate building losses (715 buildings) and the highest candidate-loss rate (86.14%), followed by Kahramanmaraş (Central) (513 buildings; 78.68%) and Defne (224 buildings; 77.24%). These findings indicate that the proposed framework is capable of identifying substantial regional differences in the spatial distribution of potential building losses.
6. Confidence-score analysis demonstrated that automatically identified candidate buildings were not limited to low-confidence detections. Mean confidence values ranged between 0.284 and 0.369 across districts, indicating that candidate losses originated from detections spanning a relatively broad confidence interval.

7. The proposed framework demonstrates that large-scale preliminary building-loss screening can be performed using only publicly accessible Google Earth imagery, without requiring auxiliary datasets such as building footprints, cadastral information, LiDAR, or digital elevation models. This characteristic makes the proposed approach particularly suitable for rapid deployment in regions where additional geospatial information is unavailable, incomplete, or difficult to obtain immediately after a major earthquake.

Overall, the proposed framework demonstrates that integrating deep-learning-based building detection with object-level temporal matching provides an efficient and scalable solution for rapid preliminary post-earthquake building-loss screening. Rather than replacing expert interpretation, the proposed methodology serves as an intelligent decision-support framework that automatically prioritizes locations requiring detailed manual inspection, thereby substantially reducing the effort associated with large-scale post-disaster damage assessment.

6 Discussion and Future Work

Numerous studies on post-earthquake building damage assessment have utilized auxiliary datasets such as building footprints, GIS layers, LiDAR data, digital surface models (DSMs), SAR imagery, and semantic segmentation outputs to improve temporal correspondence and damage estimation. Although these datasets can enhance matching accuracy, they are often unavailable, outdated, or incomplete immediately after large-scale disasters. In contrast, the proposed framework was intentionally designed to operate using only historical Google Earth imagery, thereby minimizing the dependency on additional geospatial information.

Nevertheless, several limitations should be acknowledged. Historical Google Earth imagery exhibits considerable variability in spatial resolution, viewing geometry, illumination conditions, seasonal changes, shadows, and geometric distortions, all of which increase the difficulty of both building detection and temporal matching. Furthermore, buildings that have experienced severe structural damage frequently lose their regular roof geometry, making them visually similar to debris, bare ground, or partially reconstructed structures.

Another limitation is that the current temporal matching process relies primarily on geometric correspondence. Although the proposed registration and object-matching strategy substantially improves alignment accuracy, ambiguous correspondences may still occur in densely built urban areas or regions that have undergone extensive structural changes.

Future research will focus on enhancing the temporal matching stage by incorporating additional spatial and semantic information, such as building footprints, road-network constraints, semantic segmentation outputs, transformer-based visual descriptors, and graph-based object-matching algorithms. Furthermore, integrating Google Earth imagery with very-high-resolution satellite imagery, UAV imagery, LiDAR, or SAR data, together with extending the framework from bitemporal to multi-temporal image analysis, may further improve the robustness and reliability of the proposed approach.

1735 Overall, the proposed framework provides a scalable decision-support solution for
1736 rapid preliminary building-loss screening over large earthquake-affected areas and
1737 facilitates the prioritization of locations requiring detailed expert assessment.

1738

1739 **Declarations**

1740

1741 **Funding**

1742 The author received no specific funding for this research.

1743

1744 **Competing Interests**

1745 The author declares that there are no competing interests.

1746

1747 **Author Contributions**

1748 Perihan Karakose conceived the study, designed the methodology, collected and
1749 prepared the dataset, developed the software, performed the experiments, analyzed the
1750 results, interpreted the findings, prepared the figures and tables, and wrote, reviewed,
1751 and approved the final manuscript.

1752 **Data Availability**

1753 The Satellite-II Version 1 dataset used in this study is publicly available through
1754 Roboflow Universe. The manually verified Google Earth dataset generated for this
1755 study contains historical Google Earth imagery and associated annotations. Owing
1756 to Google Earth licensing restrictions and the manually generated annotations, these
1757 data are not publicly available but may be made available by the corresponding author
1758 upon reasonable request for academic research purposes, subject to applicable licensing
1759 conditions.

1760

1761 **Code Availability**

1762 The source code developed for building detection, image registration, object-
1763 level temporal matching, and candidate building-loss analysis is available from the
1764 corresponding author upon reasonable request.

1765

1766 **Ethics Approval**

1767 Not applicable.

1768

1769 **Consent to Participate**

1770 Not applicable.

1771

1772 **Consent for Publication**

1773 Not applicable.

1774

1775

1776

1777

1778

1779

1780

References

- [1] Liu, C., Sui, H., Wang, J., Ni, Z., Ge, L.: Real-time ground-level building damage detection based on lightweight and accurate YOLOv5 using terrestrial images. *Remote Sensing* **14**(12), 2763 (2022)
- [2] Abdi, G., Esfandiari, M., Jabari, S.: Building damage detection in post-event high-resolution imagery using deep transfer learning. In: *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 531–534 (2021)
- [3] Abdi, G., Esfandiari, M., Jabari, S.: A deep transfer learning-based damage assessment on post-event very high-resolution orthophotos. *Geomatica* **75**(4), 237–250 (2022)
- [4] Ji, M., Liu, L., Buchroithner, M.: Identifying collapsed buildings using post-earthquake satellite imagery and convolutional neural networks: A case study of the 2010 haiti earthquake. *Remote Sensing* **10**(11), 1689 (2018)
- [5] Ji, M., Liu, L., Du, R., Buchroithner, M.F.: A comparative study of texture and convolutional neural network features for detecting collapsed buildings (2019)
- [6] Ji, M., Liu, L., Zhang, R., Buchroithner, M.F.: Discrimination of earthquake-induced building destruction from space using a pretrained CNN model. *Applied Sciences* **10**(2), 602 (2020)
- [7] Ghaffarian, S., Kerle, N., Pasolli, E., Jokar Arsanjani, J.: Post-disaster building database updating using automated deep learning: An integration of pre-disaster OpenStreetMap and multi-temporal satellite data. *Remote Sensing* **11**(20), 2427 (2019)
- [8] Zhan, Y., Liu, W., Maruyama, Y.: Damaged building extraction using modified Mask R-CNN model using post-event aerial images of the 2016 kumamoto earthquake (2022)
- [9] Kalantar, B., Ueda, N., Al-Najjar, H.A.H., Halin, A.A.: Assessment of convolutional neural network architectures for earthquake-induced building damage detection based on pre- and post-event orthophoto images. *Remote Sensing* **12**(21), 3529 (2020)
- [10] Li, Y., Hu, W., Dong, H., Zhang, X.: Building damage detection from post-event aerial imagery using single shot MultiBox detector. *Applied Sciences* **9**(6), 1128 (2019) <https://doi.org/10.3390/app9061128>
- [11] Ilmak, D., Iban, M.C., Şeker, D.Z.: A geospatial dataframe of collapsed buildings in antakya city after the 2023 kahramanmaraş earthquakes using object detection

based on YOLO and VHR satellite images. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3915–3919 (2024)

[12] Ali, M.U., Sultani, W., Ali, M.: Destruction from sky: Weakly supervised approach for destruction detection in satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* **162**, 115–124 (2020)

[13] Tilon, S., Nex, F., Kerle, N., Vosselman, G.: Post-disaster building damage detection from earth observation imagery using unsupervised and transferable anomaly detecting generative adversarial networks. *Remote Sensing* **12**(24), 4193 (2020)

[14] Duarte, D., Nex, F., Kerle, N., Vosselman, G.: Damage detection on building façades using multi-temporal aerial oblique imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **IV-2/W5**, 29–36 (2019) <https://doi.org/10.5194/isprs-annals-IV-2-W5-29-2019>

[15] Döşa, M.E., Seyrek, E.C., Uysal, M.: Deep learning based building change detection by integrating building footprint data to pre and post-earthquake VHR satellite images from february 6, 2023, kahramanmaraş earthquake: A case study for hatay-antakya. *International Journal of Remote Sensing* **46**(20), 7615–7638 (2025)

[16] Xia, H., *et al.*: A deep learning application for building damage assessment using ultra-high-resolution remote sensing imagery in turkey earthquake. *International Journal of Disaster Risk Science* **14**(6), 947–962 (2023)

[17] Xie, Z., Zhou, Z., He, X., Fu, Y., Gu, J., Zhang, J.: Methodology for object-level change detection in post-earthquake building damage assessment based on remote sensing images: OCD-BDA. *Remote Sensing* **16**(22), 4263 (2024)

[18] Su, S., Nawata, T., Fuse, T.: Building change detection from bitemporal aerial images using deep learning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **V-2-2020**, 565–571 (2020)

[19] Khodaverdi Zahraee, N., Rastiveis, H.: Object-oriented analysis of satellite images using artificial neural networks for post-earthquake buildings change detection. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **XLII-4/W4**, 139–144 (2017)

[20] Tong, X., *et al.*: Building-damage detection using pre- and post-seismic high-resolution satellite stereo imagery: A case study of the may 2008 wenchuan earthquake. *ISPRS Journal of Photogrammetry and Remote Sensing* **68**, 13–27 (2012)

[21] Gupta, R., *et al.*: Creating xBD: A dataset for assessing building damage from satellite imagery. In: Proceedings of the IEEE/CVF Conference on Computer

Vision and Pattern Recognition Workshops, pp. 10–17 (2019)	1873
[22] Janalipour, M., Mohammadzadeh, A.: A novel and automatic framework for producing building damage map using post-event LiDAR data. <i>International Journal of Disaster Risk Reduction</i> 39 , 101238 (2019)	1874 1875 1876 1877
[23] Zha, Y., Gao, J., Ni, S.: Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. <i>International Journal of Remote Sensing</i> 24 (3), 583–594 (2003)	1878 1879 1880 1881
[24] Alisjahbana, I., Li, J., Zhang, Y.: DeepDamageNet: A two-step deep-learning model for multi-disaster building damage segmentation and classification using satellite imagery. <i>arXiv preprint arXiv:2405.04800</i> (2024)	1882 1883 1884 1885
[25] yolo: Satellite-II Dataset. Roboflow Universe, Version 1. Open Source Dataset, CC BY 4.0. Accessed: August 1, 2026 (2024). https://universe.roboflow.com/yolo-xhpbs/satellite-ii	1886 1887 1888 1889
[26] Miura, H., Aridome, T., Matsuoka, M.: Deep learning-based identification of collapsed, non-collapsed and blue tarp-covered buildings from post-disaster aerial images (2020)	1890 1891 1892
[27] Braik, A.M., Koliou, M.: Automated building damage assessment and large-scale mapping by integrating satellite imagery, GIS, and deep learning. <i>Computer-Aided Civil and Infrastructure Engineering</i> 39 (15), 2389–2404 (2024)	1893 1894 1895 1896
[28] Wang, Y., Chew, A.W.Z., Zhang, L.: Building damage detection from satellite images after natural disasters on extremely imbalanced datasets. <i>Automation in Construction</i> 140 , 104328 (2022)	1897 1898 1899 1900
[29] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> 39 (6), 1137–1149 (2017)	1901 1902 1903 1904
[30] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: SSD: Single shot multibox detector. <i>arXiv preprint arXiv:1512.02325</i> (2015)	1905 1906 1907
[31] Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> 42 (2), 318–327 (2020)	1908 1909 1910 1911
[32] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., Ding, G.: YOLOv10: Real-time end-to-end object detection. In: <i>Advances in Neural Information Processing Systems</i> , vol. 37, pp. 107984–108011 (2024)	1912 1913 1914 1915
[33] Khanam, R., Hussain, M.: YOLOv11: An overview of the key architectural enhancements. <i>arXiv preprint arXiv:2410.17725</i> (2024)	1916 1917 1918

1919 [34] Tian, Y., Ye, Q., Doermann, D.: YOLOv12: Attention-centric real-time object
1920 detectors. In: Advances in Neural Information Processing Systems, vol. 38, pp.
1921 78433–78457 (2026)

1922

1923 [35] Bhangale, U., Durbha, S., Potnis, A., Shinde, R.: Rapid earthquake damage detec-
1924 tion using deep learning from VHR remote sensing images. In: Proceedings of the
1925 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp.
1926 2654–2657 (2019)

1927

1928 [36] Mazzanti, P., *et al.*: Assessing the performance of multi-resolution satellite SAR
1929 images for post-earthquake damage detection and mapping aimed at emergency
1930 response management. Remote Sensing **14**(9), 2210 (2022)

1931

1932 [37] Havivi, S., Rotman, S.R., Blumberg, D.G., Maman, S.: Damage assessment in
1933 rural environments following natural disasters using multi-sensor remote sensing
1934 data. Sensors **22**(24), 9998 (2022)

1935

1936 [38] Alataş, E.O., Taşkın, G.: Attribute profiles in earthquake damage identification
1937 from very high resolution post event image. In: Proceedings of the IEEE Inter-
1938 national Geoscience and Remote Sensing Symposium (IGARSS), pp. 9299–9302
1939 (2019). <https://doi.org/10.1109/IGARSS.2019.8899308>

1940

1941

1942

1943

1944

1945

1946

1947

1948

1949

1950

1951

1952

1953

1954

1955

1956

1957

1958

1959

1960

1961

1962

1963

1964