Comparative Analysis of Deep Learning Approaches for Earthquake Damage Assessment: A Case Study Incorporating Self-Supervised Contrastive Learning in the 2023 Turkish Earthquakes

Ozan Güven*, Arnaud Poletto*, Fulvia Malvido Montandon†
*Department of Computer Science, †Department of Architecture, Civil and Environmental Engineering
École Polytechnique Fédérale de Lausanne, Switzerland
{ozan.guven, arnaud.poletto, fulvia.malvidomontandon}@epfl.ch

***Abstract*—This study aims to semantically segment building damages resulting from the February 2023 earthquakes in Turkey, specifically focusing on the Kahramanmaraş region. We employ three Deep Learning methods renowned for their efficacy in change detection using pre- and post-event optical imagery. These methods include fully convolutional neural networks, employing Early Fusion and Siamese architectures. To address the challenge of limited training data for this specific task, we incorporate an inventive approach of self-supervised learning through Siamese contrastive learning to initially train the model encoders. Moreover, considering the significant class imbalance and the complexities involved in the labeling process of our dataset, we framed this task as an anomaly detection challenge. These approaches are meticulously tailored to discern subtle and significant alterations on buildings in the urban fabric caused by the seismic events, demonstrating their potential in enhancing post-disaster analysis and response.**

***Index Terms*—Earth observation, change detection, building damage detection, earthquake, fully convolutional neural network, self-supervised machine learning**

I. Introduction

The day of February 6, 2023 witnessed a catastrophic series of earthquakes that devastated southern and central Turkey and northern and western Syria, marking one of the most severe seismic events in recent history. The initial earthquake, measuring a magnitude of M_{w} 7.8 happened in the early morning of that day near the cities of Kahramanmaraş and Gaziantep [1], [2], followed approximately nine hours later by another significant quake of magnitude M_{w} 7.5, approximately 95 kilometers north-east from the first one [3], [4]. This seismic activity was accompanied by over 30 000 aftershocks, exacerbating the devastation [5].

The impact of these earthquakes was profound and far-reaching. Nearly 16 million people in Turkey and Syria were affected, with the death toll exceeding 50 000 and injuries surpassing 107 000 [6]–[8]. The earthquakes led to the displacement of more than 2 million people and caused significant destruction on infrastructure, including the destruction of over 36 000 buildings and rendering more than 311 000 buildings unusable due to their sustained damages, including vital facilities like hospitals and schools. [9]

The magnitude of this disaster necessitates an urgent and comprehensive understanding of its effects, particularly in terms of building damages. Rapid and accurate assessment of such damages is crucial for efficient disaster response and recovery efforts. It enables not only the allocation of resources and aid to the most affected areas but also supports the planning of reconstruction efforts and the mitigation of future risks.

In this context, the deployment of remote sensing technologies, especially Very High Resolution (VHR) optical images, emerges as the optimal method for analysing and assessing earthquake-induced damages. This is particularly relevant in densely populated urban settings such as Kahramanmaraş. VHR imagery, characterised by a spatial resolution finer than 10 meters, is particularly adept at detecting and quantifying the extent of damage to individual buildings and structures. This high level of detail is critical for identifying specific areas of destruction, enabling a more targeted and effective response.

Additionally, the recent advancements in VHR optical imagery have elevated its value beyond precision. Now recognised for cost-effectiveness, extensive coverage, and rapid data acquisition, these images play a crucial role in modern disaster management strategies.

Our study leverages Deep Learning techniques to analyse pre- and post-earthquake imagery from the Kahramanmaraş region, aiming to semantically segment damaged buildings. By employing Deep Learning methods renowned for change detection, we intend to contribute significantly to post-disaster analysis and response capabilities. Our approach, which integrates self-supervised learning through Siamese contrastive learning, is designed to enhance the models' ability to detect subtle and significant changes in the urban landscape resulting from the seismic events. Such methodologies not only offer insights into the scale and specifics of the damages but also hold the potential to revolutionise the way we respond to and recover from such catastrophic events.

We also present a meticulously labeled dataset based on the pre- and post-earthquake optical satellite imagery of the Kahramanmaraş region. This dataset is a compilation of high-

resolution images, both before and after the earthquake, providing a comprehensive view of the affected areas.

II. Related Works

A variety of methodologies have been employed to estimate earthquake-induced damages on buildings. Contreras et al. [10] presented a comprehensive review of data sources for building damage assessment post-earthquake, highlighting the combination of fieldwork, omnidirectional imagery, terrestrial laser scanning, and remote sensing. They emphasised the necessity of integrating multiple data sources for effective earthquake reconnaissance and the growing significance of crowdsourcing and social media platforms in data collection.

Dell'Acqua et al. [11] examined the use of optical and Synthetic Aperture Radar (SAR) data in earthquake damage assessment. They explored different contexts such as mono- and multi-temporal techniques, underscoring the potential and limitations of each approach.

Specific studies focused on the same region as our current research. Sun et al. [12] introduced the QuickQuakeBuildings dataset, combining SAR and optical data for rapid damaged-building detection. Wang et al. [13] evaluated urban building damage from the 2023 Kahramanmaraş, Turkey earthquake using SAR change detection. They combined SAR amplitude and phase coherence change detection, verifying their results against high-resolution optical images and Artificial Intelligence (AI) recognition outcomes.

In the realm of change detection, Deep Learning has emerged as a critical tool for processing data from large-scale Earth Observation (EO) systems. Particularly for segmentation tasks, the U-Net architecture, introduced by Ronneberger et al. [14], has gained significant popularity, initially in medical imagery and now extending to other domains. Utilising pre- and post-disaster images leads to advanced change detection methods. A notable contribution in this field is the work of Daudt et al. [15], who introduced three Deep Learning methods based on Fully Convolutional Neural Network (FCNN) architectures for change detection in multi-temporal Red, Green and Blue (RGB) images. Their work underscores the effectiveness of Deep Learning in this context.

Additionally, recent studies have applied Deep Learning to the specific context of the 2023 Turkey earthquakes. Malmgren and Karlbrg [16] explored the use of a dual-task U-Net Deep Learning model for automated remote damage assessment using VHR imagery. Their study highlighted the potential of Deep Learning models in building damage assessment. Similarly, Robinson et al. [17] employed AI methods to assess building damage in southeast Turkey, in collaboration with Turkey's Ministry of Interior Disaster and Emergency Management Presidency (AFAD). Their work, focusing on the immediate aftermath of the earthquake, provided crucial insights into the extent of damage across various cities, demonstrating the practical application of Deep Learning in rapid disaster response scenarios. These studies underscore the growing role of AI and Deep Learning in enhancing the accuracy and efficiency of damage assessments following major seismic events.

Building on these foundational studies, our research takes a pioneering step by integrating Deep Learning techniques to analyse earthquake-induced building damages. Our work uniquely addresses the challenges of data scarcity and class imbalance by implementing self-supervised learning approaches, particularly leveraging Siamese contrastive learning, as introduced by Mo et al. [18]. By combining the strengths of Deep Learning with the detailed data provided by VHR optical images, our study aims to make a substantial contribution to the field of earthquake damage assessment, offering a more nuanced and effective approach for post-disaster analysis and response strategies.

III. Methodology

A. Data Collection

Our study uses high-resolution optical satellite imagery from Maxar's Open Data Program [19]. The dataset comprises a total of $1\,614$ multi-temporal RGB images of the Kahramanmaraş region in Turkey, captured both before and after the February 2023 earthquakes. Each image has a resolution of $30\mathrm{cm}$ and dimensions of $17\,408 \times 17\,408$ pixels, offering a detailed view of the area's urban and rural landscapes. Examples of such images can be found in Figure 1.

B. Data Processing

Given the substantial size of the images, we extract smaller patches of $1\,024 \times 1\,024$ pixels, resulting in 289 unique patches per image. This strategy is essential to manage the computational load and to focus on specific areas within each image.

To ensure the relevance and quality of our data, we apply several filtering criteria:

- Black Pixel Threshold: We exclude images with more than 5% black pixels to eliminate areas with insufficient data.
- Snow Detection: We remove images with a mean sum of RGB pixel values below 500 to filter out snow-covered areas. We omitted snow-covered images as they can hide key features like buildings and roads, complicating accurate damage assessment post-earthquake.
- Variance Threshold: We select images with more than 4% variance in pixel values to focus on urban regions and exclude homogeneous landscapes like forests or empty spaces. This threshold was chosen through manual experimentation to effectively prioritise urban regions while excluding homogeneous landscapes such as meadows or vacant areas.

These criteria lead to a refined dataset of $26\,188$ pre- and post-event images. Further filtration based on the availability of corresponding pre- or post-event images reduces the dataset to $14\,562$ images.

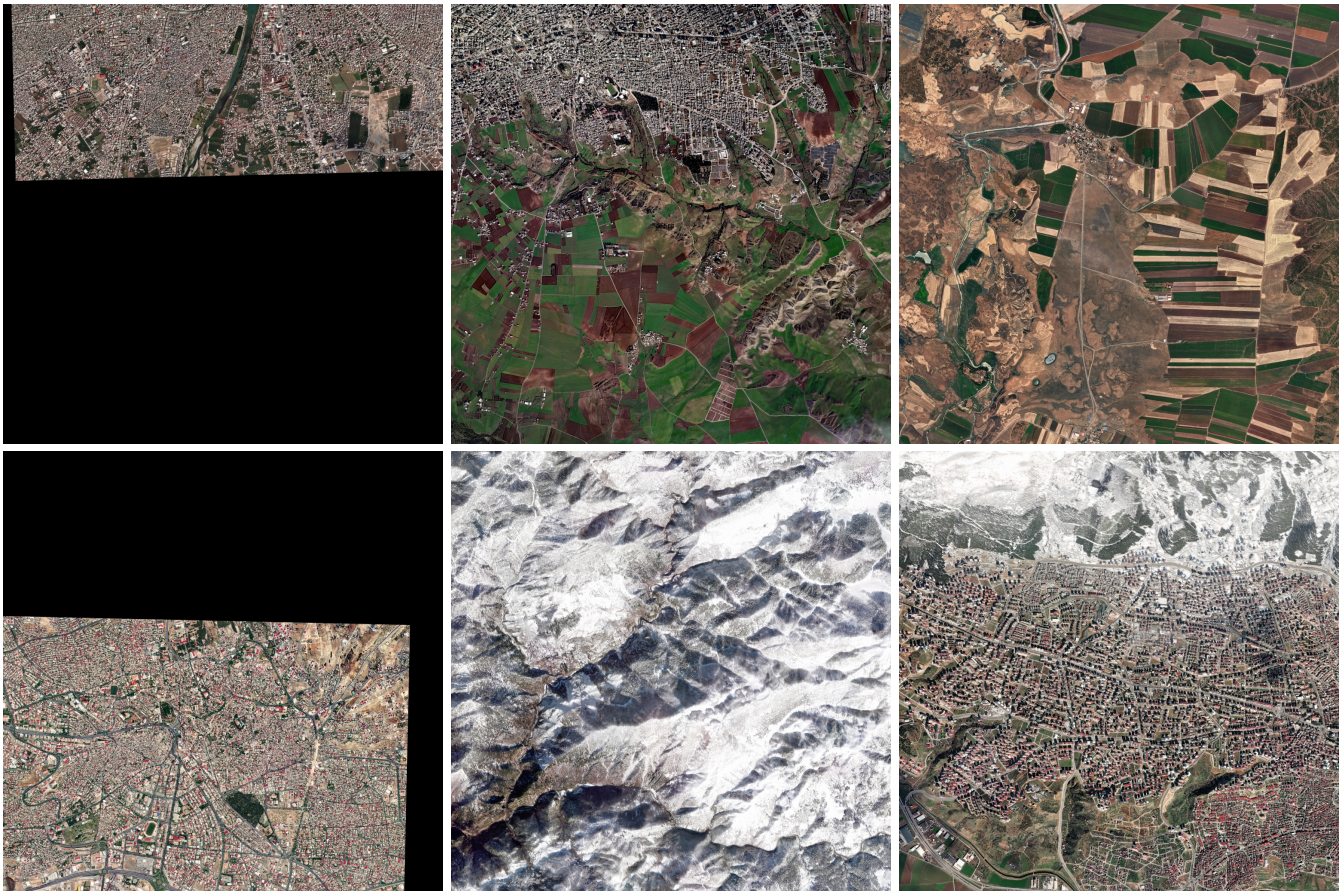


Fig. 1: Examples of various raw satellite images from Maxar.

C. Data Labelling

We identify a subset of 245 image pairs that contain visible building damage and 948 pairs without damage, based on manual inspection using a custom application designed to display pre- and post-event images side by side. Examples of pairs of images can be seen in Figure 2.

For the image pairs with detected damage, we manually create binary masks to indicate damaged areas. Our focus is on visually identifiable damage, such as collapsed structures or significant debris accumulation. Changes due to construction activities, such as new buildings, are not considered as damage. Examples of pairs of images with their corresponding masks can be found in Figure 3.

D. Data Partitioning

1) Main Training Phase: Our approach to dividing the dataset into distinct segments for training, validation, and testing involves a randomised splitting process. We allocated 70% of our dataset to the training set, ensuring a robust and comprehensive learning environment for the model. The validation and test sets each received an equal share of 15% of the data. This balanced distribution allows for effective model tuning and accurate performance evaluation.

A key aspect of our training methodology is the exclusive use of image pairs labeled as damaged for the main training phase. This strategic choice is underpinned by our objective to focus the model's learning on critical data reflecting damage scenarios. We opted not to include intact image pairs in the training set, based on the rationale that the damaged image pairs already encompass sufficient instances of undamaged structures. This is attributed to the inherent sparsity of segmentation labels within these pairs, which typically contain both damaged and undamaged areas.

2) Self-Supervision Training Strategy: In the self-supervision phase of training, we adopt a nuanced approach to dataset split, particularly for the FC-EF encoder model. Recognising the importance of contrasting damaged and undamaged scenarios in our contrastive learning framework, we include pairs of images that exhibit no damage. This inclusion is crucial for developing a robust encoder capable of discerning between damaged and intact structures. For the FC-EF encoder, we employ a dataset split ratio of 92% for training and 4% for validation and testing. This is because each training step takes 2 pair of images, squaring the space of possible data training, thereby significantly enhancing the diversity and richness of the training set, which is vital for effective contrastive learning.

For the FC-Siam models, we maintain the original split of 70% for training and 15% for validation and testing. In this

(a) Pre- and post-event images with damage.

(b) Pre- and post-event images with damage.

(c) Pre- and post-event images with no visible damage.

(d) Pre- and post-event images with no visible damage.

Fig. 2: Examples of various pre- and post-event pairs of patches after filtering.

case, each training step takes a single pair per image, thus the standard split ratio suffices as it does not necessitate the expanded possibility space needed for the FC-EF model.

E. Data Augmentation

To enhance the robustness and generalisability of our models, we have implemented an extensive data augmentation strategy. Given the relatively limited size of our dataset, data augmentation serves as a pivotal technique to simulate a more varied and expansive training environment. Our augmentation pipeline introduces a series of transformations that mimic real-world variations in lighting, color, and orientation.

These transformations include random adjustments to brightness and contrast, application of different blur effects, and the introduction of noise—all of which aim to make our models more resilient to common image quality issues. We also apply geometric augmentations such as horizontal and vertical flips, and random shifts, scales, and rotations to prepare the models for a wide range of spatial variations they may encounter in actual disaster-stricken imagery.

By diversifying the training set through these augmentations, we enable our models to learn from a broader spectrum of conditions, reducing the risk of overfitting and improving their ability to generalise from the training data to unseen images.

F. Limitations of the Approach

Our methodology, while robust, encounters certain limitations such as data interpretation challenges and labeling expertise constraints.

Optical satellite images, despite their high resolution, sometimes pose difficulties in distinguishing damaged buildings. This is due to factors such as angle discrepancies between pre- and post-event images, seasonal and meteorological variations affecting color and contrast, shadow variations, and cloud cover.

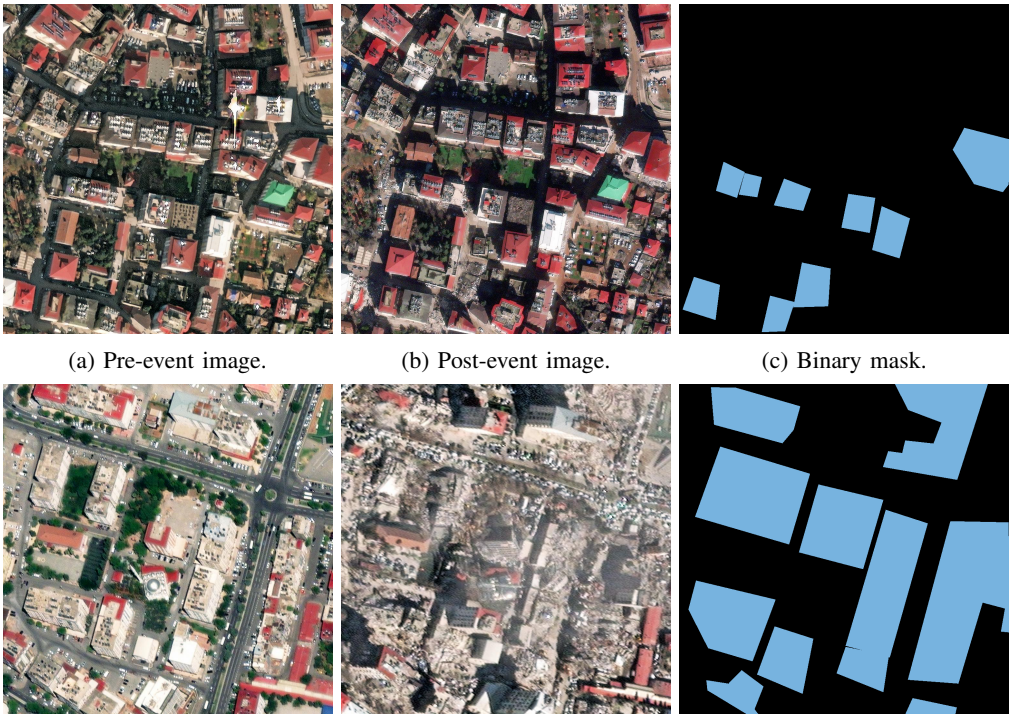
Our team, lacking in formal expertise in damage assessment, may have inadvertently overlooked subtle instances of damage or introduced bias. Additionally, due to time and resource constraints, not all potential image pairs were analysed, which might have led to under-representation of certain damage types or areas.

G. Network Architectures

In line with the naming conventions established in [15], this section details the architecture of the three networks utilised in our study.

1) Fully Convolutional Early Fusion: The first network, Fully Convolutional Early Fusion (FC-EF), draws inspiration from the U-Net architecture, featuring an encoder-decoder structure with skip connections. The distinctive feature of FC-EF is its approach to input handling: it concatenates the pre- and post-event images, resulting in a 6-channel input. This method treats the combined imagery as an extended set of color channels, providing a comprehensive view of the changes induced by the earthquake.

2) Fully Convolutional Siamese with Concatenation: The second network, Fully Convolutional Siamese with Concatenation (FC-Siam-Conc), bears resemblance to FC-EF but



(a) Pre-event image. (b) Post-event image. (c) Binary mask.

(d) Pre-event image. (e) Post-event image. (f) Binary mask.

Fig. 3: Examples of pairs of pre- and post-event images with their corresponding masks.

introduces a key variation in its structure. Instead of a single encoder, FC-Siam-Conc employs two separate encoders, each processing one of the pre- or post-event images. These encoders share the same weights, reflecting the traditional Siamese network design. The model merges the outputs of these encoders after the convolutional layers. Unique to FC-Siam-Conc is the method of combining skip connections: during the decoding phase, the skip connections from each encoding stream are concatenated, enriching the network's ability to capture and integrate the nuances of change.

3) Fully Convolutional Siamese with Difference: Finally, the third network, Fully Convolutional Siamese with Difference (FC-Siam-Diff), extends the concept introduced in FC-Siam-Conc with an innovative twist in handling skip connections. Rather than a direct concatenation, it computes the absolute value of the differences between the connections. This method underscores the network's focus on detecting distinct changes between the pre- and post-event scenarios, enhancing its sensitivity to alterations caused by the earthquake.

H. Self-Supervision for Pretraining Encoders

We integrate self-supervision as a pivotal component of our methodology, primarily to enhance the performance of our models given the complexity of the task at hand. We adopt a Siamese contrastive learning approach for pretraining the encoders. This technique revolves around the principle of spatial-temporal analysis of image pairs, categorising them based on the presence or absence of earthquake-induced damages.

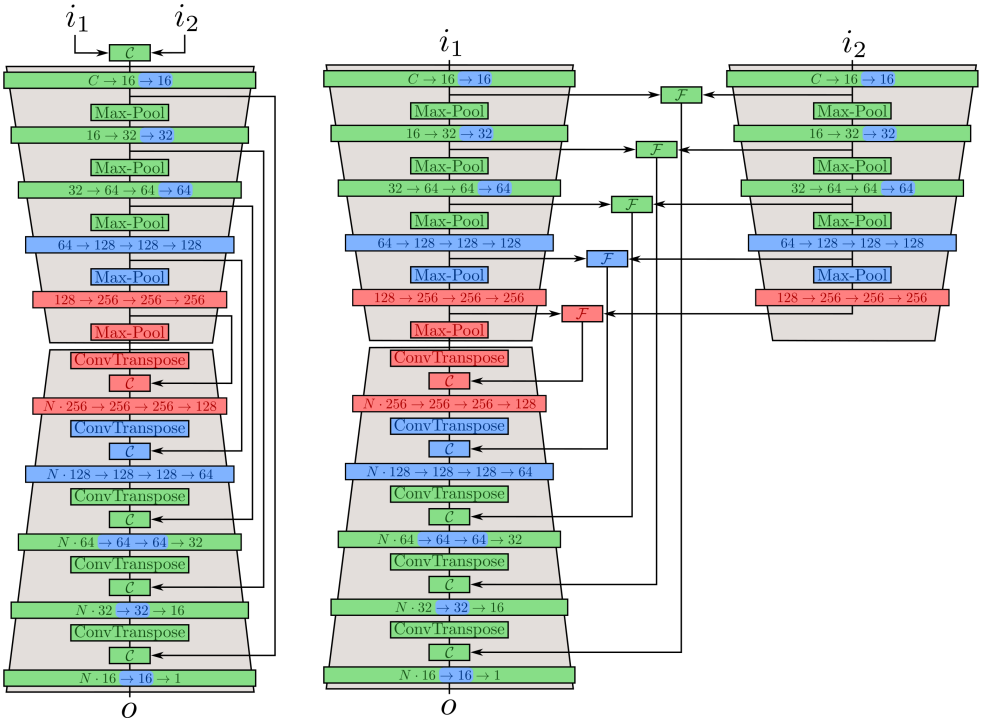
The core idea is to manipulate the embeddings generated by the encoders in a way that aligns with the damage status of the images. Specifically, for image pairs that exhibit no damage, the aim is for pre- and post-event image embeddings in the feature space to be proximate, signifying similarity. Conversely, for pairs that display damages, the embeddings should be distinctly separated in this space, indicating notable differences.

This approach is slightly modified in the case of the FC-EF network. Here, the model evaluates embeddings from two distinct pairs of images. The criterion for closeness in the embedding space is based on whether the pairs reflect the same damage status — both depicting damage or both no damages. If the pairs show differing statuses, their embeddings are expected to diverge significantly.

The mechanism of this self-supervised learning is governed by a contrastive loss function, employing the L_2 norm. This loss function is pivotal as it quantitatively measures the closeness or disparity of the embeddings, and is defined as follows:

$$\mathcal{L}(z_1, z_2) = (1 - l) \cdot \mathcal{L}_D + l \cdot [M - \mathcal{L}_D]_+^2 , \qquad (1)$$

where z_1 and z_2 represent the embeddings, l the label, $\mathcal{L}_D = ||z_1 - z_2||^2$ the squared Euclidean distance between the embeddings, and $[\cdot]_+ = max(0, \cdot)$ the element-wise maximum operation. In the case of FC-Siam-Conc or FC-Siam-Diff networks, the embeddings correspond to pre- and post-event images respectively. For the FC-EF network, they represent embeddings of two pairs of pre- and post-event images. The



(a) FC-EF model architecture.

(b) FC-Siam model architecture.

Fig. 4: FC Model Architectures: Green indicates the small model used for hyperparameter tuning, blue denotes medium-scale enhancements, and red signifies additional components for the large-scale model.

label l is assigned a value based on the damage status: it is set to 0 if there is no damage in the pair (FC-Siam-Conc or FC-Siam-Diff) or if both pairs depict the same damage status (FC-EF), and 1 if there are damages or if the pairs depict different damage statuses, respectively. The margin M is a predefined threshold that determines how far apart the embeddings of different classes should be.

I. Evaluation Metrics and Losses

To assess the performance of our models, we use various metrics, including the accuracy, precision, recall, $F1$ score and the Intersection over Union (IoU) metric.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}, \tag{5}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{6}$$

where TP represents True Positives, TN denotes True Negatives, FP stands for False Positives and FN signifies False Negatives.

We utilise three distinct loss functions to train our models, each having its own characteristics and advantages in the context of change detection.

1) Binary Cross-Entropy (BCE) Loss: This loss function is a widely used method in binary classification tasks. It is particularly effective in scenarios where we need to distinguish between two classes, such as the presence or absence of damage in change detection. The BCE loss is mathematically defined as:

$$\mathcal{L}_{BCE} = -\left[y \log(p) + (1 - y) \log(1 - p)\right], \tag{7}$$

where y is the binary label (0 or 1) and p is the predicted probability of the positive class.

2) IoU Loss: Also known as the Jaccard loss, this function is particularly useful for assessing the overlap between the predicted and actual damaged areas. The IoU loss is given by:

$$\mathcal{L}_{IoU} = 1 - \frac{\sum_i p_i y_i}{\sum_i (p_i + y_i) - \sum_i p_i y_i}, \tag{8}$$

where y_i is the binary label for pixel i and p_i is the predicted probability of the positive class for pixel i.

3) Dice Loss: Closely related to the $F1$ score, the Dice loss is particularly effective in handling class imbalance, which is a common issue in change detection tasks. It is defined as:

$$\mathcal{L}_{Dice} = 1 - \frac{2 \cdot \sum_i p_i y_i}{\sum_i p_i + \sum_i y_i} \tag{9}$$

Similar to the $F1$ score, it balances the precision and recall, making it suitable for scenarios where both false positives and false negatives are equally costly.

Each of these loss functions brings a unique perspective to the learning process, making them suitable for different aspects of change detection.

IV. Experimental Setup

A. Network Architectures

Figure 4 depicts the comprehensive architecture of the models. The encoder in all models operates through a series of convolutional layers, each followed by batch normalisation, dropout, and ReLU activation, denoted by a single arrow ($\rightarrow$) in the figure. The final layer of these models deviates from this pattern by only applying the convolutional layers and directly outputting logits instead. Downsampling occurs at regular intervals via max pooling until the image is reduced to a compact representation. The FC-EF and FC-Siam models differ initially in their channel input: $C = 3$ for FC-Siam, whereas $C = 6$ for FC-EF, a result of the early-stage image concatenation in FC-EF by the module $\mathcal{C}$. In the subsequent upsampling phase, transposed convolutional layers expand the embeddings, which are then combined with skip connections. For FC-EF, these connections are simply prior layer embeddings, but for the FC-Siam models, the fusion module $\mathcal{F}$ actively integrates them—employing absolute difference for FC-Siam-Diff and depthwise concatenation for FC-Siam-Conc. The parameter N, within the decoder framework, represents the count of distinct data streams that are integrated at each layer. For FC-EF and FC-Siam-Diff architectures, $N = 2$ indicates the integration of two streams: the upsampled feature maps and the corresponding skip connections. Conversely, in the FC-Siam-Conc architecture, $N = 3$ reflects the merger of three streams at each decoder layer, which includes the upsampled feature maps along with the skip connections from both encoders.

B. Hyperparameter Tuning

We implement an extensive hyperparameter tuning process for the three Deep Learning models, employing a random search strategy. This method involves systematically exploring a diverse array of hyperparameter combinations. The hyperparameters under consideration include encoder and decoder layer sizes, accumulation steps, dropout rates, learning rates, loss functions, and weight decay parameters. We conducted this tuning separately for each model to identify the optimal configuration for our specific task.

We aim to maximise the validation IoU metric, which is critical for assessing the performance of our models.

Table I outlines the hyperparameters and their respective distribution values used in our hyperparameter tuning process. In Figures 4a and 4b, you can find descriptions of the models, along with the respective differences between the small, medium, and large versions.

To provide a clear understanding of the specific hyperparameter configurations selected for each model post-tuning, we have detailed the final settings in Table II. This table presents the optimised hyperparameters for the three models.

C. Main Training Strategy

Our training process was meticulously structured for all three methods, each method being trained not only in its standard form but also in two variations incorporating self-supervised pre-training. The training was conducted over a duration of 100 epochs, utilising the AdamW optimiser [20], [21]. Our training strategy focuses on optimising the IoU metric. This metric was selected due to its proficiency in addressing data imbalance.

The training was conducted on a high-performance Nvidia RTX 3080 GPU equipped with $10\mathrm{GB}$ of memory. Each epoch was observed to take approximately $15\mathrm{s}$, with the dataset being divided into roughly 33 batches. This calculation translates to an overall training duration of about 25 minutes for the 100 epochs.

Crucially, our training regimen included running the three primary models alongside six additional variants. These variants were specifically designed to assess the impact of our self-supervised pre-training approach on the overall model performance. We experimented with two key configurations for each variant: one where the encoder was pre-trained without freezing its parameters, allowing further adjustments during the main training phase; and another where the encoder was pre-trained and then frozen, preventing any further modifications during subsequent training. This comprehensive approach allowed us to thoroughly evaluate the effectiveness of self-supervised pre-training in enhancing model performance.

D. Self-Supervised Training Strategy

In line with our approach outlined in Section III-H, self-supervised training primarily focuses on refining the encoder component. An essential preparatory step involved adapting the encoders for optimal processing of the final convolutional outputs. This adaptation entails the introduction of an average pooling layer, which averages over the depths of the convolution map. The resulting output is then flattened and channeled through a single linear layer. This layer yields an embedding of 128 dimensions, creating a compact yet expressive representation that facilitates easy comparison between different embeddings. This process is carefully designed to minimise the introduction of learnable parameters that could potentially lead to overfitting, thereby ensuring that the burden of representation learning predominantly rests on the encoder.

For the auxiliary contrastive task, our training strategy harnesses a balanced mix of damaged and intact data pairs. In each training iteration, we randomly sample a pair of images (or two pairs in the case of the FC-EF model) from the dataset. This method maintains an equilibrium between positive and negative comparisons within the training batches. The training parameters are meticulously chosen: we utilise a batch size of 8, a learning rate of 10^{-5}, and a weight decay of 10^{-4}. Additionally, we do not employ dropout in this phase and set a contrastive margin of 1.0.

Hyperparameter	Distribution Type	Values
Loss	Categorical	BCE, Dice, IoU
Batch Size	Categorical	4, 8, 16, 32
Learning Rate	Log Uniform	10^{-5} to 10^{-2}
Weight Decay	Log Uniform	10^{-5} to 10^{-1}
Dropout Rate	Uniform	0 to 0.5
Channels	Categorical	Small, Medium, Large

TABLE I: Hyperparameters tuned for each model, with their respective distribution values.

Hyperparameter	Model		
	FC-EF	FC-Siam-Conc	FC-Siam-Diff
Loss	Dice	IoU	Dice
Batch Size	4	4	8
Learning Rate	$4 \cdot 10^{-5}$	$5 \cdot 10^{-4}$	$3 \cdot 10^{-4}$
Weight Decay	$7 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
Dropout Rate	0.01	0.02	0.05
Channels	Large	Large	Large

TABLE II: Optimised hyperparameter settings for each of the three models.

Given the disparity in the number of single images relative to image pairs, the training duration for each model is adjusted accordingly. The FC-EF model undergoes training for 1 epoch, encompassing 121 000 potential combinations of pair of image pairs. Conversely, for both FC-Siam models, the training extends over 200 epochs, each epoch comprising 268 possible combinations of image pairs.

This self-supervised training phase was executed using the same hardware setup as the main training procedure. Both training phases spanned approximately 2h, providing an extensive and detailed learning period for the encoders to effectively adapt and refine their capabilities.

V. Results

The outcomes of the comparative evaluation are presented in Table III. Additionally, Figure 5 provides a selection of predictions generated by the FC-Siam-Diff model, which has not undergone pretraining.

The model demonstrates swift inferencing capabilities, requiring approximately 2.6s to process an evaluation set on the same hardware specifications detailed previously. The testing dataset consists of 28 images, each with dimensions of 1024×1024 pixels.

VI. Discussion

A. Comparative Analysis of Model Performance

In this section, we delve into the comparative analysis of our three proposed models, focusing on their performance metrics as outlined in the initial segment of Table III.

1) Accuracy of Base Models: Upon examining the base models without a pre-trained encoder, we observe that all three models – FC-EF, FC-Siam-Conc and FC-Siam-Diff – demonstrate high accuracy, exceeding 94%. This high accuracy is anticipated, considering the unbalanced nature of our dataset, where most pixel values are 0 (indicating no damage), with a minority being 1 (indicating damage).

2) Advantages of Siamese Models: A noteworthy observation is the enhanced performance of the Siamese models (FC-Siam-Conc and FC-Siam-Diff) compared to the traditional FC-EF model across all metrics. In particular, FC-Siam-Diff exhibits a notable edge in precision, achieving a score of 59.12%. It also secures the top IoU score, indicating its superior ability to identify damaged areas accurately. When it comes to the $F1$ score, which is a measure of overlap between the predicted and actual damaged areas, FC-Siam-Diff again stands out, achieving 55.62%. This high $F1$ score underscores the model's effectiveness in accurately delineating the areas impacted by the earthquake.

3) Overall Model Efficacy: The comparative analysis highlights the superior performance of the Siamese-based models, especially FC-Siam-Diff, in terms of precision, $F1$ score, and IoU score. These models, with their unique architectures, demonstrate enhanced capabilities in detecting and delineating earthquake-induced damages in the Kahramanmaraş region, offering valuable tools for effective disaster response and recovery planning.

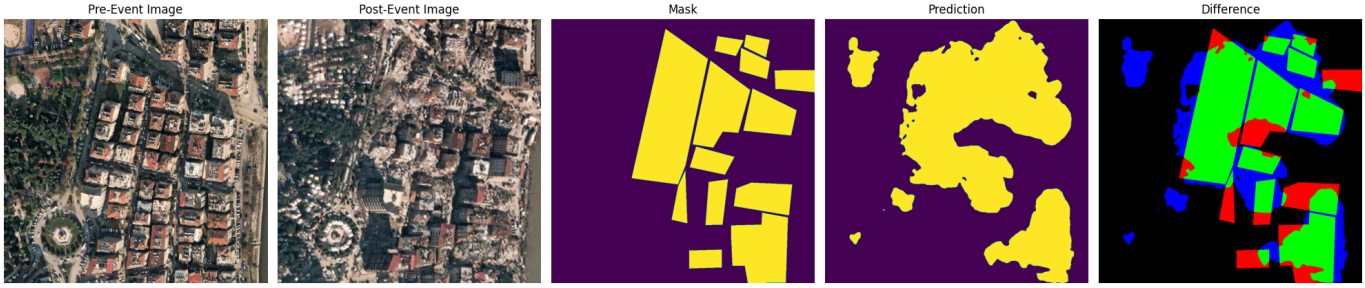
B. Self-Supervised Learning Impact

In this section, we examine the effects of self-supervised pretraining on our models, particularly focusing on how it influences the recall and overall performance.

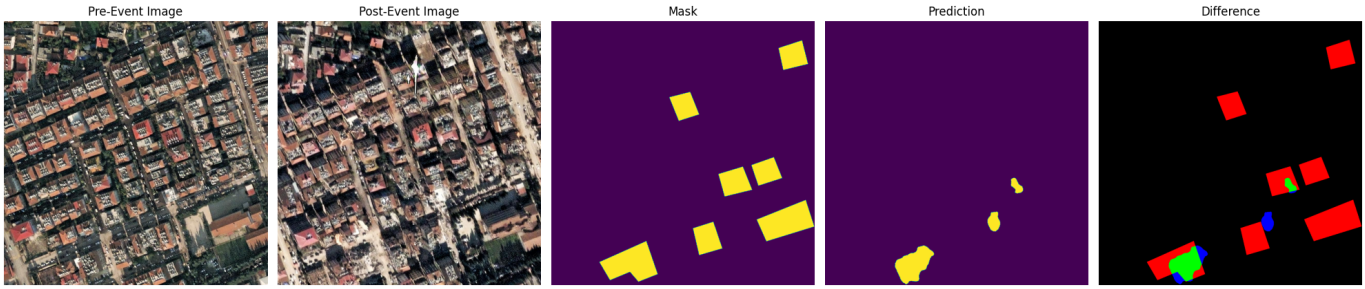
1) Enhanced Recall with Pretrained Weights: Our findings indicate improvement in recall for models with pretrained weights. Specifically, the FC-Siam-Conc model shows an approximate 4% increase in recall, while the FC-Siam-Diff model exhibits a 6% boost. This suggests that models preloaded with pretrained weights are more proficient in identifying a higher fraction of actual positive cases (i.e., damaged areas). This aspect is particularly relevant to our study, as it aligns with our objective of anomaly detection in earthquake-affected regions.

2) Comparison with Base Models: When assessing the overall effectiveness of our models, which were pretrained

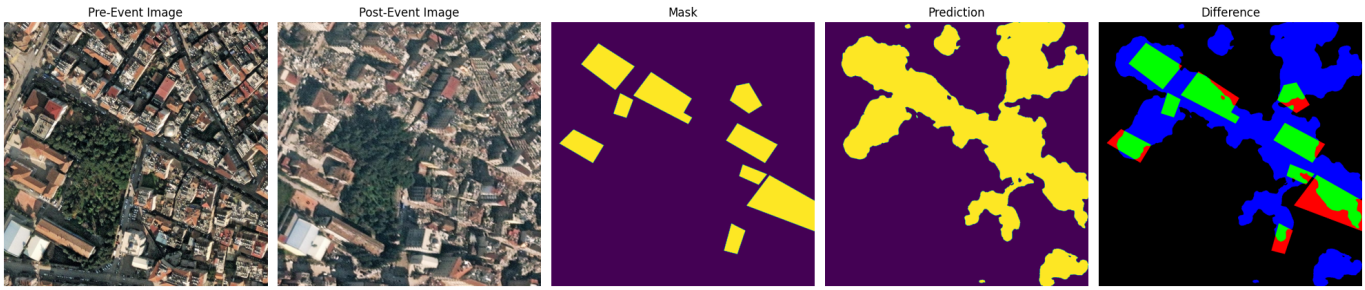
(a) A scene with extensive earthquake-induced changes; the model accurately identifies major building damages.



(b) An example showcasing the model's proficiency in damage detection, with a high degree of accuracy across the image.



(c) An instance of suboptimal prediction, where the model fails to identify several damaged buildings.



(d) A case of overestimation, where the model erroneously marks extensive areas as damaged.

Fig. 5: Illustrative examples of pre- and post-earthquake imagery, alongside their respective damage masks, predictions from the FC-Siam-Diff model with non-pretrained encoder, and a comparison of predicted versus actual damage masks. In the comparison, green denotes true positives, red represents false negatives, and blue signifies false positives.

Model	Pretrained Encoder	Frozen Encoder	Accuracy (%)	Precision (%)	Recall (%)	$F1$ (%)	IoU (%)
FC-EF	✗	✗	94.45	53.54	53.15	53.34	36.37
FC-Siam-Conc			94.58	54.12	55.95	55.02	37.95
FC-Siam-Diff			**94.99**	**59.12**	52.50	**55.62**	**38.52**
FC-EF	✓	✗	94.76	57.46	47.21	51.83	34.98
FC-Siam-Conc			93.30	45.09	**59.83**	51.42	34.61
FC-Siam-Diff			93.20	44.79	58.87	50.87	34.11
FC-EF	✓	✓	88.80	26.64	49.82	34.71	21.00
FC-Siam-Conc			92.99	41.69	45.87	43.68	27.94
FC-Siam-Diff			88.84	25.47	44.95	32.52	19.42

TABLE III: Results of the three models across the different methods and metrics evaluated on the test set.

using comparative analysis, we found that their performance is not better than the baseline models. This indicates that, although the models are capable of learning from the auxiliary task, this learning does not necessarily enhance their performance in subsequent segmentation tasks. Several reasons could explain this observation:

- **Task Generalisation Challenges:** The auxiliary task's focus on creating distinct embeddings for pre- and post-event images may not align closely enough with the requirements of the segmentation task. Although the model learns, this might not be adequately applicable to the segmentation task.
- **Suboptimal Freezing of Encoder Weights:** Our results where the encoder was completely frozen showed a significant drop in performance, underlining the counter-productive nature of this approach. Given the dissimilar nature of the auxiliary and primary tasks, it seems more beneficial to allow the model's encoder to continue learning to maintain and build upon its acquired knowledge.
- **Uniform Learning Rate Across Encoder and Decoder:** Applying the same learning rate to both encoder and decoder might lead to catastrophic forgetting. This is evidenced in our case, as most of the learning results were slightly worse than the baseline, indicating that the knowledge gained during the supervised phase might have been overridden in the subsequent learning phase, leading to no significant improvement. To mitigate this, a smaller learning rate for the encoder, as opposed to the decoder, could be employed to prevent such forgetting and preserve the beneficial aspects of the initial training.

These findings reveal that models without a pretrained encoder generally yield the best scores in terms of overall performance metrics. However, when it comes to recall, a crucial measure for our study, models with a pretrained encoder demonstrate a significant advantage. This enhancement in recall is particularly relevant in the context of accurately identifying damaged areas post-earthquake, where minimising false negatives is important. Therefore, while the overall performance is essential, the superior recall of pretrained models aligns more closely with our primary objective of effectively and accurately detecting earthquake-induced damages.

C. Case Study Analysis

The set of images provided offer a detailed visual account of the FC-Siam-Diff model's performance in the context of earthquake damage assessment.

In the initial example, as seen in Figure 5a, we observe a segment of the urban area that has experienced considerable damage due to seismic events. The model adeptly identifies areas of structural damage, such as compromised buildings. Notably, the stadium, visible in the post-event image with tents for those displaced by the quake, contrasts starkly with its pre-event state of a simple grassy sports field. Impressively, our model does not mistakenly classify this as a damaged area, demonstrating its capacity to discern between actual structural damage and changes in the environment used for disaster response.

The second example in Figure 5b continues to demonstrate the model's strengths, where it remains resilient against alterations in traffic and shifts in perspective between image captures. The model's predictions exhibit high concordance with the verified damaged zones, suggesting a strong performance in scenarios where damages are visually distinct and well-defined.

Conversely, in the third instance depicted in Figure 5c, the model does not perform as well. It overlooks several damaged structures, which brings to light the challenges of detecting subtle or partially concealed damages. This instance emphasises the necessity for models to evolve beyond recognising explicit destruction and to develop an acute sensitivity to the subtler indicators of damage.

Finally, in the example shown in Figure 5d, the model displays a tendency to overpredict, incorrectly marking large areas as damaged. This type of overconfidence could be attributed to the model's overreaction to certain patterns or textures it wrongly associates with damage, culminating in a heightened false positive rate. Such examples underscore the importance of fine-tuning the model to balance sensitivity with specificity, ensuring accurate damage assessment in post-disaster scenarios.

VII. Limitations

In the realm of damage assessment, particularly in our study, we approach the task through a binary segmentation lens. This methodology, while effective in distinguishing damaged

from undamaged areas, does not capture the varied degrees of damage. Such granularity is critical for accurately directing aid effectively and estimating costs. The complexity of damage, ranging from minor to catastrophic, necessitates a more nuanced scale. Standardisation attempts have already been made, like in the work of Ehrlich et al. [22]. Adopting a similar damage assessment scale, one that encapsulates the spectrum of destruction, would mark a significant enhancement.

Our training dataset, while foundational to our model's development, presents its own set of limitations. The dataset's scope and diversity are constrained by our predefined criteria for labeling damage. While this approach ensures consistency, it inadvertently introduces a degree of bias. This stems from a singular perspective on what constitutes damage, potentially overlooking variations that other experts or locals might perceive. To mitigate this, embracing a more collaborative approach to data collection and labeling is essential. Crowdsourcing emerges as a powerful tool in this context, harnessing insights from a broad spectrum of contributors, including experts, practitioners, and volunteers. The development and deployment of a Python application for streamlined damage assessment marks a leap forward in efficiency. However, expanding this framework to an online platform or integrating it with existing datasets could exponentially enrich our data pool. Such an expansion would not only diversify our dataset but also enhance the robustness and generalisability of our models.

Lastly, external factors such as weather conditions, including seasonal changes that can significantly alter the appearance of fields and landscapes, lighting variations, shadows, and changes in viewing angles pose significant challenges to our models' accuracy. These environmental and situational variances can drastically alter the appearance of landscapes and structures, complicating the task of consistent damage detection. For instance, a building that appears undamaged under one lighting condition might reveal significant destruction under another. Similarly, snow cover, shadowing, or seasonal vegetation changes can mask or mimic signs of damage. Additionally, construction sites or landfills might be mistaken for destroyed areas due to their similar appearance. Tackling these challenges requires adaptive and resilient models, capable of discerning true changes from mere visual discrepancies. In EO, various methods have been developed to address such issues, ranging from preprocessing techniques that normalise lighting conditions to algorithms designed to compensate for angle variations. Integrating these existing methods into our approach could substantially improve our models' ability to accurately interpret and assess damage under diverse environmental conditions, making them more reliable and effective in real-world scenarios.

VIII. Conclusion

This study demonstrates the application of FCNN and Siamese architectures in the task of earthquake damage assessment, with a particular focus on the 2023 Turkish Earthquakes. Through the integration of self-supervised contrastive learning, we aimed at enhancing the encoders' ability to distinguish between damaged and undamaged structures, potentially leading to a more accurate and nuanced understanding of the affected areas.

Our findings have significant implications for EO and the broader field of disaster response. The ability to rapidly and accurately assess damage post-disaster is crucial for directing aid and reconstruction efforts, and the methodologies we have presented show promise in improving these processes.

Future research could explore several avenues to build upon our work. The incorporation of Attention Mechanisms may provide a way to further refine the focus of our models on relevant features within the data. Building outlines could be used to enhance the localisation of damage, while Open Vocabulary Detection methods like Grounding DINO [23] could broaden the types of damage the models can recognise. Furthermore, the application of Transfer Learning, specifically utilising pretrained models from extensive datasets like ImageNet [24], could be tailored to fine-tune our models to accelerate the learning process and improve model robustness.

The potential improvements to our approach include exploring the use of larger datasets to improve generalisability, and incorporating additional modalities of data, such as SAR or thermal imagery. Future work should also consider the development of models that can operate across a wider range of conditions and are robust to the environmental changes that can affect visual data, such as lighting and seasonal variations.

In conclusion, our study contributes a valuable perspective to the use of Deep Learning in EO for disaster response, particularly in the context of earthquake damage assessment. As EO technology continues to advance, we anticipate our work will play a role in shaping the development of more sophisticated, accurate, and efficient tools for disaster assessment and management.

References

[1] B. Üniversitesi Kandilli Rasathanesi ve Deprem Araştırma Enstitüsü Bölgesel Deprem-Tsunami İzleme ve Değerlendirme Merkezi, "06 Şubat 2023 Sofalca Şehitkamil Gaziantep Depremi," http://www.koeri.boun.edu.tr/sismo/2/06-subat-2023-ml7-4-sofalaca-sehitkamil-gaziantep-depremi/, 2023, accessed: 23.12.2023.

[2] U. S. G. Survey, "M 7.8 - Pazarcik earthquake, Kahramanmaras earthquake sequence," https://earthquake.usgs.gov/earthquakes/eventpage/us6000jllz/executive, 2023, accessed: 23.12.2023.

[3] B. Üniversitesi Kandilli Rasathanesi ve Deprem Araştırma Enstitüsü Bölgesel Deprem-Tsunami İzleme ve Değerlendirme Merkezi, "06 Şubat 2023 Ekinözü Kahramanmaraş Depremi," http://www.koeri.boun.edu.tr/sismo/2/06-subat-2023-ml7-5-ekinozu-kahramanmaras-depremi/, 2023, accessed: 23.12.2023.

[4] U. S. G. Survey, "M 7.5 - Elbistan earthquake, Kahramanmaras earthquake sequence," https://earthquake.usgs.gov/earthquakes/eventpage/us6000jlqa/executive, 2023, accessed: 23.12.2023.

[5] W. H. Organization, "Kahramanmaraş Earthquakes – Türkiye and Syria," https://reliefweb.int/report/turkiye/kahramanmaras-earthquakes-turkiye-and-syria-31-may-2023, 2023, accessed: 23.12.2023.

[6] U. N. O. for the Coordination of Humanitarian Affairs (OCHA), "Türkiye Earthquakes Flash Appeal," https://www.unocha.org/publications/report/turkiye/flash-appeal-turkiye-earthquake-february-may-2023-entr, 2023, accessed: 27.12.2023.

[7] ——, "Türkiye Earthquake 2023 Humanitarian Response Overview," https://reliefweb.int/report/turkiye/turkiye-earthquake-2023-humanitarian-response-overview-30-june-2023-entr, 2023, accessed: 27.12.2023.

[8] W. H. Organization, "Earthquake response in Türkiye and Whole of Syria," https://www.who.int/publications/m/item/who-flash-appeal--earthquake-response-in-t-rkiye-and-whole-of-syria, 2023, accessed: 27.12.2023.

[9] I. B. C. Relief and D. Foundation, "Devastating Earthquakes in Southern Türkiye and Northern Syria," https://reliefweb.int/report/turkiye/devastating-earthquakes-southern-turkiye-and-northern-syria-december-15th-2023-situation-report-30-entr, 2023, accessed: 23.12.2023.

[10] D. Contreras, S. Wilkinson, and P. James, "Earthquake Reconnaissance Data Sources, a Literature Review," *Earth*, vol. 2, no. 4, pp. 1006–1037, 2021. [Online]. Available: https://www.mdpi.com/2673-4834/2/4/60

[11] F. Dell'Acqua and P. Gamba, "Remote Sensing and Earthquake Damage Assessment: Experiences, Limits, and Perspectives," *Proceedings of the IEEE*, vol. 100, no. 10, pp. 2876–2890, 2012.

[12] Y. Sun, Y. Wang, and M. Eineder, "QuickQuakeBuildings: Post-earthquake SAR-Optical Dataset for Quick Damaged-building Detection," 2023.

[13] X. Wang, G. Feng, L. He, Q. An, Z. Xiong, H. Lu, W. Wang, N. Li, Y. Zhao, Y. Wang, and Y. Wang, "Evaluating Urban Building Damage of 2023 Kahramanmaras, Turkey Earthquake Sequence Using SAR Change Detection," *Sensors*, vol. 23, no. 14, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/14/6342

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015.

[15] R. C. Daudt, B. L. Saux, and A. Boulch, "Fully Convolutional Siamese Networks for Change Detection," *CoRR*, vol. abs/1810.08462, 2018. [Online]. Available: http://arxiv.org/abs/1810.08462

[16] T. Karlbrg and J. Malmgren, "Deep Learning for Building Damage Assessment of the 2023 Turkey Earthquakes : A comparison of two remote sensing methods," Ph.D. dissertation, KTH, Geoinformatics, 2023. [Online]. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-335734

[17] C. Robinson, R. Gupta, S. F. Nsutezo, E. Pound, A. Ortiz, M. Rosa, K. White, R. Dodhia, A. Zolli, C. Birge, and J. L. Ferres, "Turkey Building Damage Assessment," *Microsoft AI for Good*, February 16 2023.

[18] S. Mo, Z. Sun, and C. Li, "Siamese Prototypical Contrastive Learning," 2022.

[19] Maxar, "Turkey and Syria Earthquake 2023," https://www.maxar.com/open-data/turkey-earthquake-2023, 2023.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[21] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2019.

[22] D. Ehrlich, A. Gerhardinger, C. MacDonald, M. Pesaresi, I. Caravaggi, and C. Louvrier, "Standardized Damage Assessment and Reporting: A Methodology that Combines Field Data and Satellite Data Analysis for Improved Damage Reporting," EUR Report 22223 EN, 2006.

[23] "Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection, author=Shilong Liu and Zhaoyang Zeng and Tianhe Ren and Feng Li and Hao Zhang and Jie Yang and Chunyuan Li and Jianwei Yang and Hang Su and Jun Zhu and Lei Zhang," 2023.

[24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.