CMPT733 TERM PROJECT

Disaster damage assessment using satellite images

Navjot Kaur* | Ali Mahdavi-Amiri

School of Computing Science, Simon Fraser University, British Columbia, Canada

***Correspondence**
Ali Mahdavi-Amiri, School of Computing Science, Simon Fraser University
Email: nka77@sfu.ca

Summary

A fast and effective damage assessment helps in Humanitarian Assistance and Disaster Response (HADR) and satellite images provide an instrumental and high-coverage information to assist in such analysis. The deep learning methods have been applied to pre-disaster and post-disaster satellite images to learn and predict the damage impact on buildings. Most recent works use either a Siamese network or a difference module in the spatial domain to capture the damage and apply attention on the temporal features independently instead of their difference. Here, we propose a novel transformer based network for building damage assessment which leverages hierarchical spatial features of multiple resolutions and captures temporal difference in feature domain after applying transformer encoder on the spatial features. The proposed network achieves state of the art performance on a large-scale disaster damage dataset (xBD) for building localization and damage classification. We also evaluate the model for building structure change detection on LEVIR-CD dataset where it gives significant improvement over previous counterparts.

KEYWORDS:
Damage classification, Change detection, Vision transformers, satellite imagery, Computer vision, building segmentation

1 | INTRODUCTION

With the increasing implications of climate change, natural disasters like hurricanes and fires have been impacting thousands of lives every year. It is important to have efficient damage assessment approaches to help in Humanitarian Assistance and Disaster Response (HADR) and provide quick and accurate damage information for emergency responses and resource allocations. Damage assessments done by ground crews can be time-consuming and labor-intensive (Spencer, Hoskere, & Narazaki, 2019). To expedite building damage assessments, several studies have implemented computer vision techniques on high-resolution aerial imagery (Cheng, Behzadan, & Noshadravan, 2021; Fujita et al., 2017) and satellite imagery (Cao & Choe, 2020; McCarthy et al., 2020; Tong et al., 2012). While aerial imagery can capture more details of buildings due to lower flying altitudes, it requires extensive local planning for building detection and has smaller building coverage than satellite imagery. Satellite imagery, which provides the near real-time and high-coverage information, offers opportunities to assist in large-scale post-disaster building damage assessments (Corbane, Carrion, Lemoine, & Broglia, 2011). By leveraging satellite imagery and deep learning, the process of damage assessments can be accelerated with the generation of high-quality building footprints and damage level classification for each building. In this study, we present a novel deep learning technique to perform damage classification and building segmentation tasks. Also, we provide a high-resolution satellite imagery dataset from Hurricane Ida in 2021 for domain adaptation.

The conventional approach of implementing computer vision for building damage assessment using satellite imagery is to pose the problem as a combination of segmentation and classification task and train deep learning models on pre-disaster and post-disaster satellite images. Many researchers

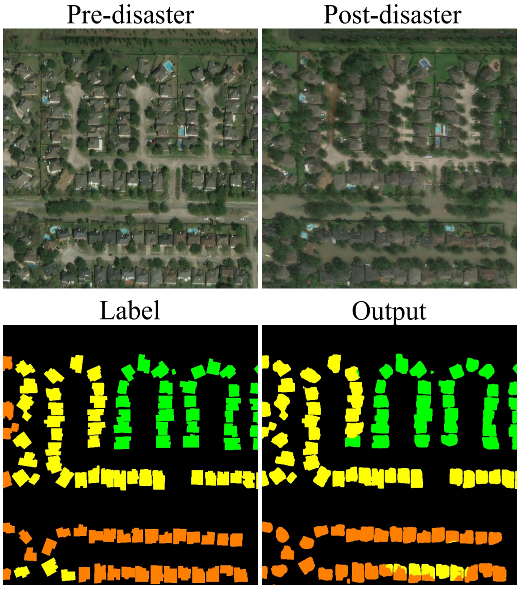


FIGURE 1 Samples from xBD dataset. The first row represents pre-disaster and post-disaster image. The second row has the ground truth label and the model prediction. The color assignment is as orange:major damage, yellow: minor damage and green:no damage

have utilized convolutional neural networks (CNNs) and achieved a significant progress.For example, (Weber & Kané, 2020) considered building damage assessment as a semantic segmentation task, in which damage levels are assigned to different class labels and pre- and post-disaster images are concatenated as input.

The problem of damage classification can be also modelled as change detection where the change is detected over a period of time with a pair of pre- and post-disaster images. The problem is however more challenging due to multi-class classification instead of binary classification and highly varying visual impact of disasters for different geographical regions. Standard practice is to use a deep Convolutional Neural Network (CNN) with a two-branch architecture utilizing feature fusion schemes from pre- and post-disaster images. For instance, Hao et al. (2021) and Wu, Zhang, Xia, Xu, & Li (2021) concatenated the features from pre- and post-disaster images and fed them into a Siamese network architecture where networks first independently learn the features from pre-disaster and post-disaster images and later a classification head compares both the feature sets. Gupta & Shah (2021) developed a framework that uses the difference between pre- and post- disaster features as input of the deep neural network, which is denoted as RescueNet. More recently, attention mechanism has been used with deep neural networks for damage classification (Shen et al., 2022) which uses cross-directional attention and data augmentation by cutting and mixing data, to model the relation between pre- and post-disaster images. On the other hand, the state of art methods for change detection heavily use attention and difference modules in a binary classification setting. Chen, Qi, & Shi (2022) and Bandara & Patel (2022) make use of attention and transformers on the encoded features followed by a difference module to efficiently learn the temporal change.

The key challenge of the damage classification problem is to efficiently capture the difference in the buildings condition. Most of the existing methods use a difference module at the end of model pipeline (in spatial domain) which makes it difficult for model to optimize for comparison. Others use concatenation or simple difference on the encoded features where change information is not fully captured. Secondly, the attention based methods (Chen et al., 2022) use the attention independently on the individual features from different temporal domains and not on the difference of the features. Another problem we observe is that these models build the output using low-resolution features and hence significant information is lost while generating a high resolution damage classification mask.

To address the above observations, we introduce a new model which is motivated by the strengths of visual transformer and UNet architecture and adapt them to change detection setting. We use UNet as a backbone to provide the initial encoding features. We further encode these features using transformers and take the difference in the feature domain and then pass it to the transformer decoders to map the features back in spatial domain. We then hierarchically build the output mask by up-sampling and concatenating the features from lower dimension to higher dimension layers in the decoder. The network has been tested on a large-scale dataset for damage classification (xBD) and on a change detection dataset (LEVIR-CD) and it outperforms the existing frameworks, as discussed in 5. We also introduce a new dataset Ida-BD with 88 image pairs (1024x1024) with a very high resolution (0.5m/pixel) from Hurricane Ida 2021. Since the near real-time damage assessment face the scarcity of labelled datasets, this dataset can be used to develop methods for domain adaptation from larger datasets like xBD. We provide a baseline for the domain adaptation task using a simple transfer learning method.

Below are the contributions of this work:

- We present a coupled UNet architecture which uses transformer based difference to perform well on both the damage classification and building segmentation tasks.
- To learn the change effectively, we apply attention on the difference of the transformer encodings in the feature domain.

- We hierarchically build the output damage mask by upsampling and concatenating the low dimension features with higher dimension features.
- We introduce a new dataset, Ida-BD, and provide a baseline for damage assessment task.

Rest of the paper is organized as follows. We first discuss the related works in the field of building damage classification and change detection using satellite images in Section 2. In Section 3 we detail our model architecture along with the loss functions and training settings. The Section 4 provides an overview of the datasets used for evaluation and the results are presented in the Section 5. Finally the last Section 6 provides the conclusions.

2 | RELATED WORK

We look into the recent methods used for building damage classification using satellite images. Since the xview2 challenge (Gupta et al., 2019) held in 2019, many researchers have tried different techniques varying from Siamese networks, fully convolutional networks and attention mechanism to solve the problem of damage classification. Since the damage classification is closely related to the change detection problem, we later look into related works for change detection in remote sensing. Many works like (Hao et al., 2021) have combined the U-Net model with the Siamese architecture where the U-Net model learns the semantic segmentation of buildings and the Siamese model focuses on the damage scale classification. (Wu et al., 2021) added attention on incoming layers from the encoder before fusing with the upsampled features. Such networks are limited by the use of separate stages for detecting buildings and damage classification. This setting makes it difficult for the model to benefit from multi-task supervision since each stage prioritize a single problem.

(Gupta & Shah, 2021) drew insights from DeepLabv3+ and adapt it for the task of damage classification by augmenting it with the capability of change detection across multi-temporal images. It uses a backbone of Dilated ResNet and make use of Atrous Spatial Pyramid Pooling module at different stride values along with image pooling to generate multi-scale features. First a segmentation head makes predictions independently for pre- and post- disaster images and then a change Detection head applies Convolutional layers along with batch normalization and ReLU activations to the difference of post- and pre-disaster features. (Weber & Kané, 2020) used a Mask R-CNN backbone augmented with a Feature Pyramid Network module and a semantic segmentation head. They exploit the multi-temporal information to predict both building locations and damage level with the same network. We note that these architecture rely on the global features learnt for pre- and post-disaster images which are then merged to learn the change (here damage) at the final stage. Since the changes are of different resolutions, a hierarchical analysis would be useful to generate a better quality change mask.

Recently, (Shen et al., 2022) have used a two stages framework for building segmentation and damage assessment. In stage 1, a single U-Net branch is used for building segmentation which uses only pre-disaster images as input and produces segmentation masks of building objects. In stage 2, the pre- and post-disaster images are fed into the two network branches separately. To further enhance the feature representations, a cross-directional fusion model is used which intends to recalibarate features from spatial and channel dimensions. Although the information is being exchanged between the two networks, the model is not explicitly enforced to learn the change between the inputs.

Mapping the features from pre- and post-disaster images to the feature domain help the model to learn the discrepancies in the spatial domain, which often occur due to imagery being captured at different daytime and date. There has not been much work on this for the damage detection task, so we review works developed for change detection on satellite images since the problem of damage detection from satellite images is closely related with the change detection (from the perspective of viewing level of damage as change). One of the works for change detection, Bi-temporal image Transformer (BiT)(Chen et al., 2022), introduced a transformer encoder to model contexts in the compact token-based space-time. A CNN backbone followed by spatial attention is used to convert each temporal image into a compact set of semantic tokens. Then a transformer encoder is used to model the context within the two context-rich token sets which are re-projected to the pixel-space by a Siamese transformer decoder followed by shallow CNN to produce pixel-level change predictions. Since the performance of transformer encoder is limited by the resolution of the visual tokens produced by CNN backbone and the classification mask is generated by direct upsampling of less than 1/4th of final resolution, the generated output is of low quality for high-resolution images. Another recent work by (Bandara & Patel, 2022) unifies hierarchically structured transformer encoder with Multi-Layer Perception (MLP) decoder in a Siamese network architecture to efficiently render multi-scale long-range details. Although the model gives good performance on the change detection tasks, the model is quite slow to train, possibly due to MLP layers.

3 | METHOD

We propose a new model architecture for multi-class change detection, based on UNet and transformers, as shown in Fig.2 .

Here we discuss the proposed model architecture and loss functions to be used, followed by methods for domain adaptation on unseen small satellite image datasets.

3.1 | Model architecture

Our model is motivated by the strengths of visual transformer and UNet architecture which have been established for effectively learning context and image segmentation tasks respectively. We first train UNet for building segmentation task on the xBD data using pre-disaster images and intermittently giving the post-disaster images. This later is used as a backbone to provide the initial encoding features. We further encode these features using transformers and take the difference in the feature domain and then pass it to the transformer decoders to map the features back in spatial domain. We then hierarchically build the output mask by upsampling and concatenating the features from lower dimension to higher dimension layers in the decoder.

Considering the damage classification as a change detection problem, the pre-disaster and post-disaster images are given as input independently to a UNet encoder. Here the UNet is pre-trained for building localization task using ResNet or DenseNet encoding module. The encoded layers from the UNet encoders with pre- and post- images are then further passed into transformer encoder which transformer densely models the context. Next absolute difference is calculated in feature domain between the outputs of transformer encoder from pre- and post- features and fed into transformer decoders to map back into spatial domain. In the decoding stage, at each up-sampling level of UNet, the transformer decoder output is added with up-sampled features, followed by convolutional layers to avoid any artifacts. This leads to hierarchical development of the damage output mask which performs well on both the classification and segmentation tasks.

3.1.1 | Loss function

We use a class weighted loss function 1 with a combination of focal loss and dice loss in order to perform well for the classification task on the unbalanced dataset. Total five classes are considered: background (0), no damage (1), minor damage (2), major damage (3) and destroyed (4) and weights w_i are assigned as per the pixel distribution of the classes in xBD dataset 1 . The focal loss L_{focal_i} is used to further address the class imbalance where a modulating term is applied to the cross entropy loss function to focus the learning on hard misclassified examples. The second loss is dice loss L_{dice_i} which tries to maximize the overlap between predicted and ground truth boundaries. The denominator considers the total number of boundary pixels at global scale while the numerator considers the intersection, implicitly capturing local behaviour.

$$
\begin{aligned}
Loss &= \sum_i w_i(L_{focal_i} + \alpha L_{dice_i}) && i \in \{0, \dots 4\} \\
L_{focal_i} &= \sum_i -(1-p_i)^{\gamma} \log(p_i) && i \in \{0, 1\} \\
L_{dice_i} &= 1 - \frac{2 \mid \tilde{c} \cup c \mid}{\mid \tilde{c} \mid \cap \mid c \mid} && c \in \{0, 1\}
\end{aligned}
\tag{1}
$$

TABLE 1 Class-wise pixels count distribution in xBD dataset and the corresponding weights used in training. The notation used: 0-Background, 1-No damage, 2-Minor damage, 3-Major damage, 4-Destroyed

	0	1	2	3	4
Pixels count	96	2.7	0.1	0.1	0.1
Weight assigned	0.01	0.1	0.7	0.7	0.7

3.1.2 | Training settings

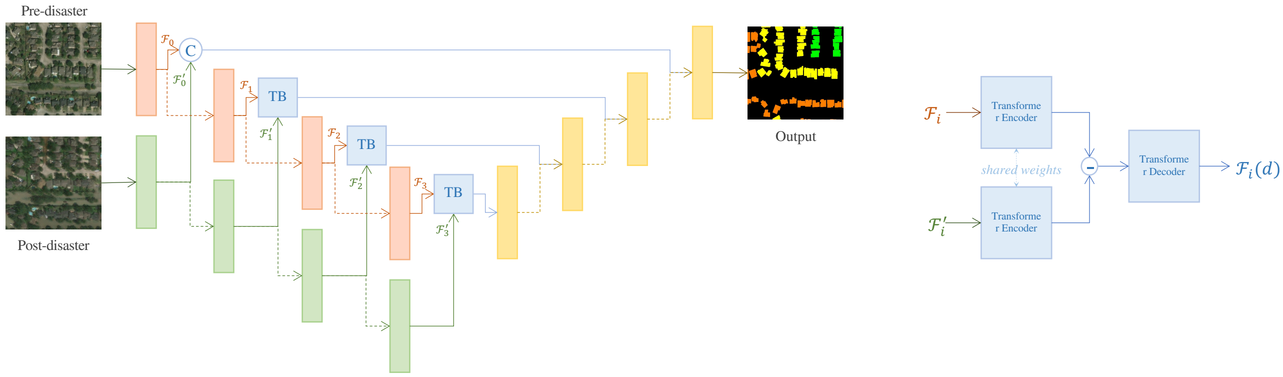
The model is trained using Adam optimizer with learning rate starting with value of 1e-4 which is gradually reduced by factor of 0.6 following multi-step learning rate scheduler. For the xBD dataset, the model takes in 1024x1024 images as input and is trained using batch size of 8 for 50 epochs. For the LEVIR dataset, model takes in 256x256 cropped images and is trained using batch size of 16 for 200 epochs. Geometric and photometric data augmentations are heavily used to make the model robust to variations and noise in the dataset. We hold 10% of the data for validation and train on the remaining set.

3.1.3 | Domain adaptation

In order to use the model on small datasets like Ida-BD 4.2, we use simple transfer learning approach to fine-tune our model on the new dataset. We run the pre-trained model from xBD dataset for 10 epochs using a fixed learning rate of 1e-6.

4 | DATASETS

The xBD dataset Gupta et al. (2019) is used for model training and testing for disaster damage detection and classification task. Later, we introduce a new dataset of higher resolution (but smaller size) from Hurricane Ida for damage assessment and explore the domain adaptation from xBD to this data. Additionally, we also use LEVIR dataset to evaluate the model for change detection task.



(a) Model Architecture

(b) Transformer block

FIGURE 2 The model architecture for damage detection and classification. A pair of pre-disaster and post-disaster images is fed into an encoder made up of stacked convolution blocks. The output from each block of pre- and post- encoders is passed jointly into transformer block which maps the spatial features into feature domain and then take the difference followed by transformer decoder to map the features back in spatial domain. Next the output mask is hierarchically built by up-sampling and concatenating the fea- tures from lower dimension to higher dimension layers.

4.1 | xBD

We use xBD (Gupta et al., 2019), which is a large-scale public dataset for building segmentation and damage assessment to benchmark our method. It consists of building segmentation and damage assessment notations for high-resolution (0.8m/pixel) satellite imagery, collected before and after 19 disaster events (such as floods, volcanic eruptions, earthquakes, and hurricanes) spread across the world. The dataset consists of geo-registered pairs of pre- and post-disaster images of size 1024 pixels x 1024 pixels with building polygons and 4-class damage labels provided for each building, as shown in Fig.1 . It uses polygons to represent building segments and provides 4 damage categories which include no damage, minor damage, major damage and destroyed for each building. We use the train split of xBD to train and test our method as the test set annotations are not publicly available yet. The training data is divided by the original authors into Tier1 and Tier3 data, with each tier correspond to a different set of disaster events. We split off about 10% of the Tier1 data into a validation set using stratified sampling across disaster events to ensure a representative sample. Our models are trained on the Tier 1 and Tier3 Train set and tested on the Tier1 validation set.

4.2 | Ida-BD

We created and made open a new dataset, Ida-BD, which has a very high resolution (VHR, 0.5m/pixel) 88 image pairs with a size of 1024 x 1024 pixels. The satellite imagery was collected close to New Orleans city in Louisiana, USA, one of the most impacted areas during Hurricane Ida in 2021. Similar to xBD, we use polygons to represent building segments and provide 4 damage categories. All annotations were done by the in-house team with quality control procedures using Labelbox (Labelbox, 2022). We first annotated building polygons in pre-disaster images to avoid incorrect building boundaries due to damage. Then, we classified building damage for each building based on post-disaster images and all annotations are reviewed by senior members in the team. Due to the small size of data, it poses a challenge of domain adaptation and we leverage the pre-trained model on xBD dataset for training and evaluation on this dataset. Since xBD dataset includes various disaster types, the class distribution of xBD dataset is quite different from Ida-BD dataset. The comparison of damage class distribution in xBD and IdaBD datasets is shown in Table 2 . To address this shift in target distribution, we update the class weights in the loss function according to the new distribution. We also use aggressive data augmentation for scaling and normalization to help model transfer knowledge from the dataset with different resolution and RGB distribution.

TABLE 2 Class-wise pixels count distribution in xBD dataset and Ida-BD dataset. The notation used: 0-Background, 1-No damage, 2-Minor damage, 3-Major damage, 4-Destroyed

	0	1	2	3	3
xBD	96.1	2.7	0.1	0.1	0.1
IdaBD	81.7	11.9	4.6	1.6	0.05

4.3 | LEVIR-CD

LEVIR-CD consists of 637 very high-resolution (VHR, 0.5m/pixel) Google Earth image patch pairs with a size of 1024 × 1024 pixels. These bi-temporal images with time span of 5 to 14 years have significant land-use changes, especially the construction growth. LEVIR-CD covers various types of buildings, such as villa residences, tall apartments, small garages and large warehouses. The change is attributed to factors such as building growth or construction and the building decline. This dataset provides binary masks as labels with values 1 for change and 0 for unchanged. The fully annotated LEVIR-CD contains a total of 31,333 individual change-building instances.

5 | EVALUATION

In this section we present the evaluation results of the model on the listed datasets in 4. We first share the metrics used for the evaluation and then present the qualitative and quantitative results. Later we discuss few ablation studies to provide a view into the impact of various components of our model and different loss functions.

5.1 | Metrics

We report results using the XView2 Challenge metrics which is a weighted average of the building segmentation F1 score and the harmonic mean of class wise damage classification F1 scores:

$$Score = 0.3 * F1_{Loc} + 0.7 * F1_{Class} \tag{2}$$

$$F1_{Loc} = \frac{2 \mid X \cup Y \mid}{\mid X \mid \cap \mid Y \mid}$$

$$F1_{Class} = \frac{1}{\frac{1}{F1_i} + \cdots + \frac{1}{F1_n}}$$

Here, $F1_{Loc}$ is the F1-Score for building segmentation and X and Y denotes background and buildings interchangeably. $F1_i$ denotes class specific F1-scores for each damage level. The metric is very challenging because it heavily penalizes overfitting to over-represented classes and the xBD dataset is heavily skewed.

5.2 | Quantitative and qualitative results

Here we discuss the results for the three tasks - damage detection and classification, change detection and domain adaptation tasks. We evaluate the results for these tasks on xBD dataset, LEVIR dataset and Ida-BD dataset respectively.

5.2.1 | Results for damage classification

We test the model on xBD dataset for damage detection and classification task by comparing the F1-score and IOU metrics. We compare the results with Siamese UNet Wu et al. (2021) which serves as a baseline for most of the works in damage detection. We compare with Dual-HRNet Ku, Seo, & Jeon (2020) which was one of the top-performing architectures in xview2-challenge Gupta et al. (2019). Additionally, we compare with the numbers published by RescueNet Gupta & Shah (2021) and BDANet Shen et al. (2022) which are attention based models. We observe that our model performs well on all the metrics, as shared in 3 . Considering the class-wise performance individually, BDANet performs better for 'major' and 'destroyed' class. However we note that our model performs well in all the classes jointly and gives higher overall F1-score and IOU. The qualitative results are displayed in Fig.3 where our model assigns the same damage level to all pixels within a building boundary.

5.2.2 | Results for change detection

We evaluate the performance for change detection on LEVIR dataset and compare it against the recent published works based on transformers Chen et al. (2022) Bandara & Patel (2022). Our model performs significantly better in both overall F1-score and IOU metrics, as shared in 4 . Additionally, we observe that adding a convolution layer after every merge of outputs from transformer decoder and lower layer, provides smoother results and less artifacts. The accuracy achieved under this setting is slightly lower than our final model, though higher than other recent works. The qualitative results are displayed in Fig.4 where we can see that the model produces finer building boundaries and capture few additional buildings, which had been missed by other models.

5.2.3 | Results for domain adaptation

To evaluate on Ida-BD dataset, we experimented with three different training settings for our model and use Siamese-UNet as a baseline model. Firstly, we train the model on Ida-BD data from scratch and test it. In the second setting, we initialize the model with weights from pre-trained model on xBD data and compare its performance. Lastly we also evaluate the pre-trained model directly, without any fine-tuning on the Ida-BD dataset. Since the percentage of 'destroyed' category is relatively lower than rest of the classes 2 , we merge the 'destroyed' with 'major damage' class and hence frame the problem as 4-class classification for the fine-tuning. As shown in the table 5 , the best performance is obtained by initializing the model with weights from the network trained on xBD data and then fine-tuning on the Ida-BD dataset. The pre-training

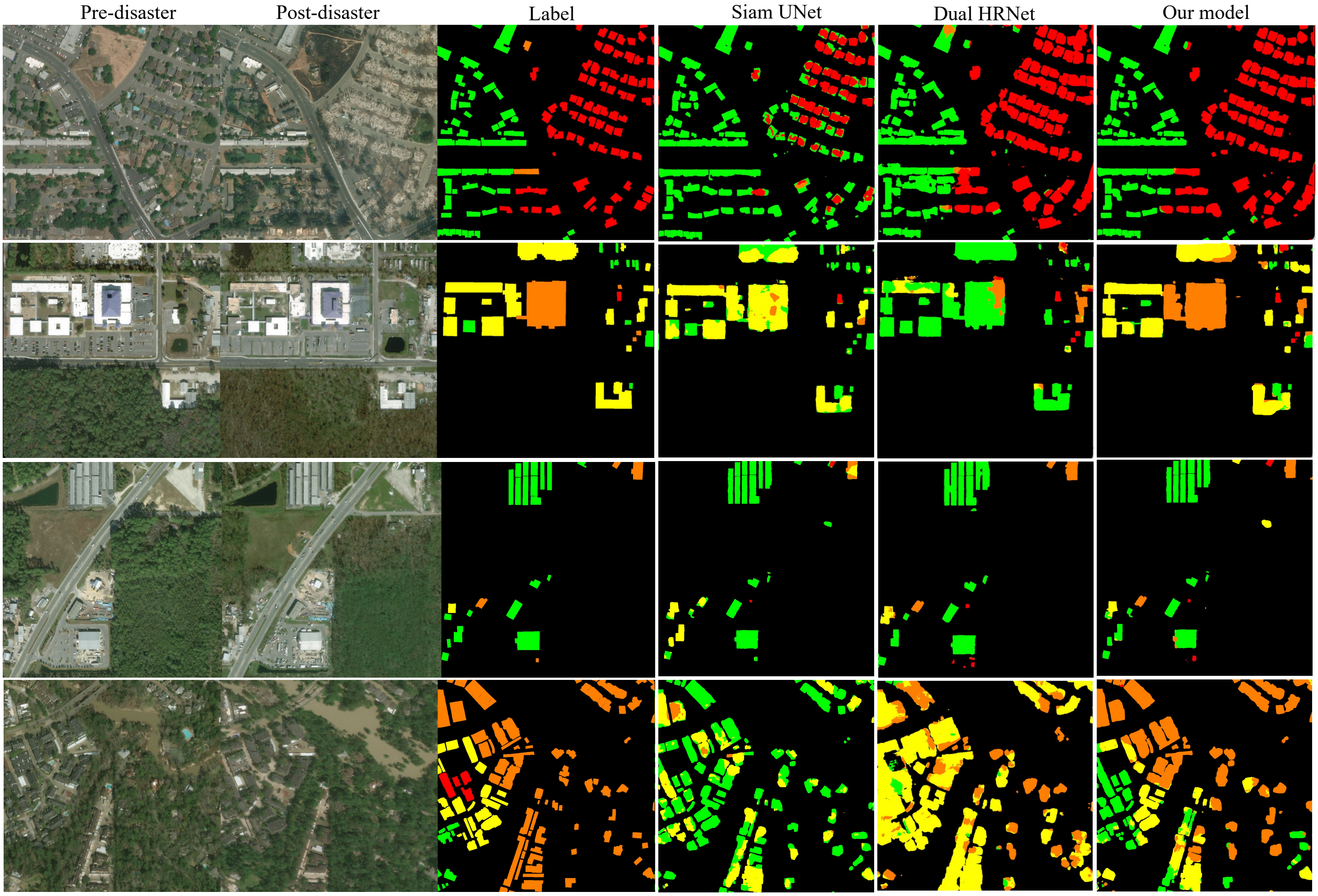


FIGURE 3 Qualitative results for Damage Classification (evaluation on xBD dataset)

TABLE 3 Average Quantitative Results for Damage Classification: xBD dataset

				Class F1-scores			
Model	**Score**	**IOU**	**F1-Score**	**No Damage**	**Minor Damage**	**Major Damage**	**Destroyed**
Siam-UNet	0.743	0.824	0.709	0.955	0.576	0.744	0.662
Dual-HRNet	0.769	0.834	0.741	0.898	0.590	0.737	0.809
RescueNet[a]	0.766	0.840	0.735	0.883	0.563	0.771	0.808
BDANet[a]	0.806	0.864	0.782	0.925	0.616	**0.788**	**0.876**
Ours	**0.819**	**0.872**	**0.796**	**0.978**	**0.711**	0.765	0.772

on the large-scale xBD dataset provides the model a good prior and the model is able to generalize better. The qualitative results are displayed in the fig. 5 .

5.3 | Ablation studies

Below we discuss few experiments on the impact of different loss functions and various components of our model.

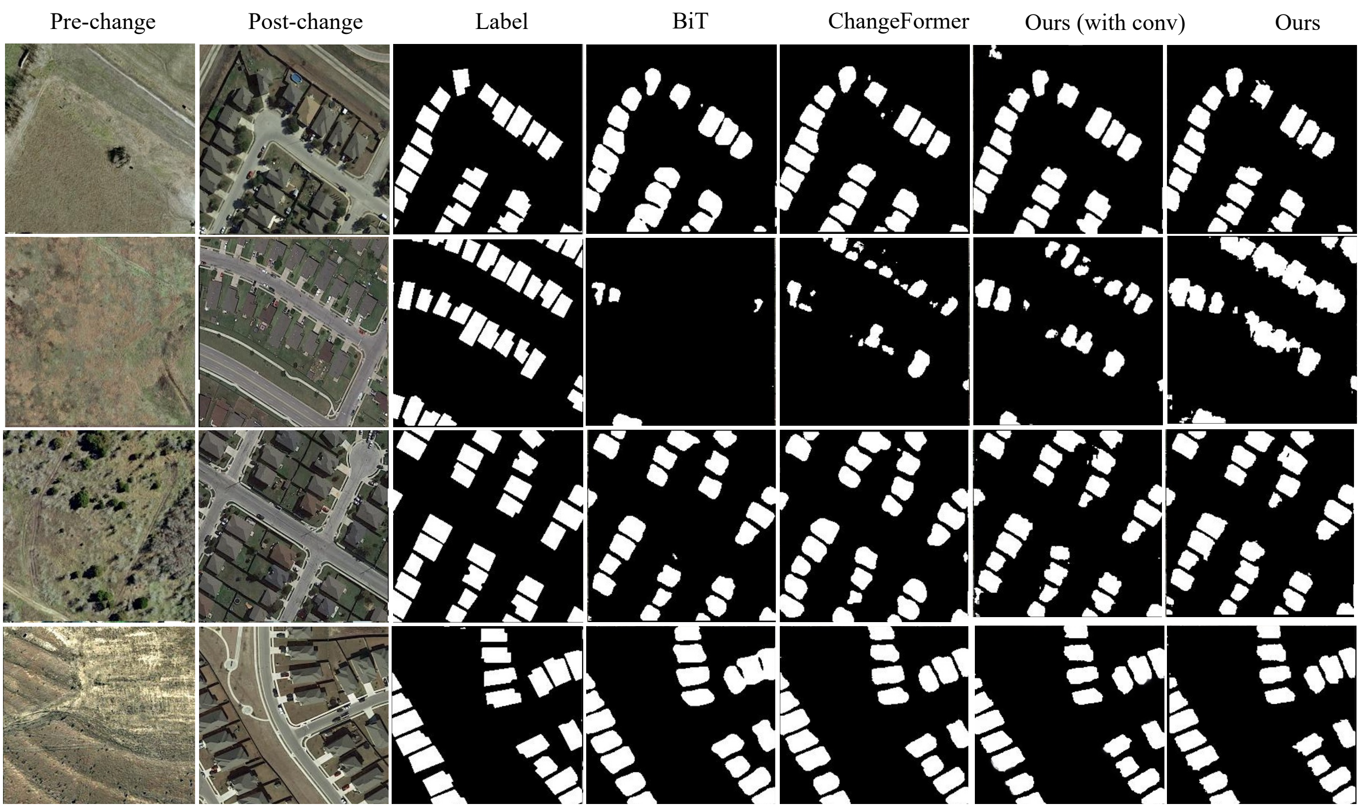


FIGURE 4 Qualitative results for Change Detection (evaluation on LEVIR-CD dataset)

TABLE 4 Average Quantitative Results for Change Detection: LEVIR-CD dataset

			Class F1-scores	
Model	**IOU**	**F1-Score**	**0 class**	**1 class**
Siam-UNet	0.813	0.859	0.987	0.788
BiT	0.829	0.899	0.990	0.807
Changeformer	0.828	0.898	0.990	0.806
Ours + with conv	0.833	0.901	0.991	0.812
Ours	**0.842**	**0.908**	**0.991**	**0.825**

TABLE 5 Domain Adaptation on Ida-BD dataset

					Class F1-scores		
Model	**Pre-training**	**Training**	**IOU**	**F1-Score**	**No Damage**	**Minor Damage**	**Major Damage**
Siam-UNet	-	Ida	0.697	0.472	0.906	0.313	0.483
	xBD	Ida	0.748	0.507	0.846	0.322	0.609
	xBD	-	0.584	0.307	0.916	0.208	0.251
Ours	-	Ida	0.778	0.541	**0.916**	0.384	0.538
	xBD	Ida	**0.805**	**0.585**	0.910	**0.439**	**0.577**
	xBD	-	0.674	0.023	0.881	0.617	0.008

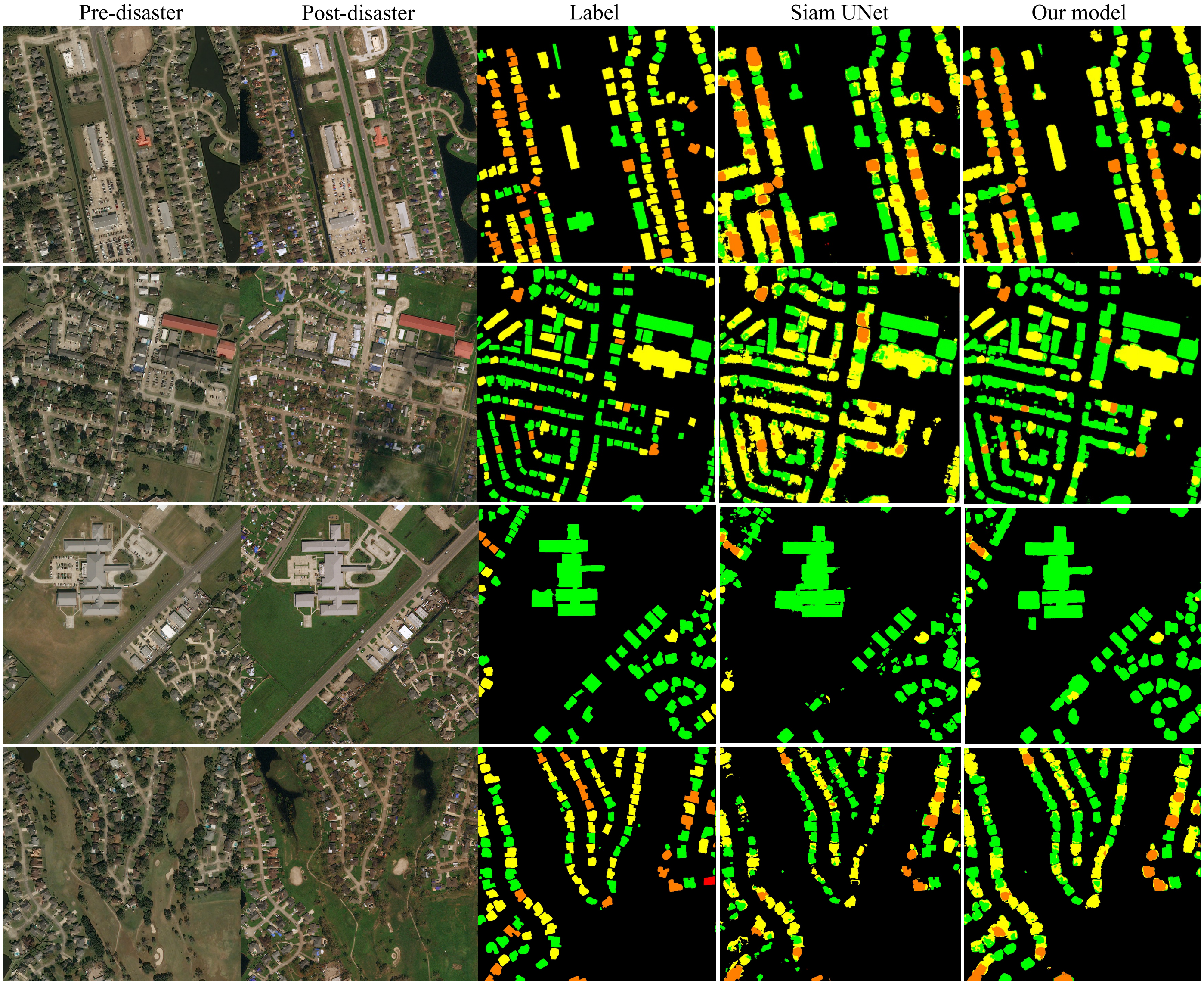


FIGURE 5 Qualitative results for Domain Adaptation (evaluation on Ida-BD dataset)

5.3.1 | Loss functions

We experimented with different combinations of loss functions which are presented in the table 6 . The localization aware loss (Gupta & Shah, 2021) gave us the better F1-scores for damage classes where pixel is accounted in the cross-entropy loss calculation only in the presence of buildings. However, the segmentation results were relatively poor and we had to depend on segmentation mask generated from pre-trained UNet model. We also tried ordinal loss in the form of mean square error between ground truth and predicted output for only buildings, by converting the values from scale of 1-4 to the scale of 0-1. This loss did not work well for our task which could be because of the highly imbalanced and sparse class distribution. The weighted sum of focal and dice loss provided the best

TABLE 6 Impact of Loss functions (xBD data)

Loss function	IOU	F1-Score
Focal + Dice	**0.816**	0.796
Focal + Dice + Ordinal	0.807	0.791
Buildings only Cross-entropy	0.798	**0.803**

TABLE 7 Effect of the number of Transformer layers used (LEVIR data)

Number	IOU	F1-Score
0	0.818	0.843
1	0.837	0.897
2	**0.843**	0.903
3	0.842	**0.908**
4	0.838	0.901

performance in class wise F1 scores as well as building boundary precision, since the former explicitly addresses the class imbalance and the latter ensures crisp building boundaries.

5.3.2 | Transformer layers

We experimented with adding different number of transformer blocks between encoder and decoder and three gave us the best results. We observe that adding transformer blocks at all levels, except the highest resolution layer, provide an increment in accuracy. The addition in final layer does not perform well which could be because the input dimension is very high and it becomes hard for transformers to learn in this space. Hence we use only convolutional layer in the highest resolution layer. On the other hand, if the transformer is not used in any of the layers, the performance highly degrades, indicating the importance of transformer encoded features. The test result values from LEVIR-CD dataset are shared in the table 7 . Also, using a combination of transformer encoder and decoder along with the difference module, instead of just transformer encoder followed by difference module, provided a bump of accuracy by 0.8% for xBD dataset.

6 | DISCUSSIONS AND CONCLUSIONS

The model developed in this study can effectively detect damaged buildings, as well as their damage levels, by using high-resolution satellite imagery. The building damage detection results generated by the model provide better understanding of disaster impacts within an area through large-scale investigations by using satellite imagery. In addition, the domain adaptation results indicate that the model can be adapted to a new event with only little fine-tuning, which is critical for the application of the model for future events. The results of this model can inform emergency responders and decision-makers of the locations of highly impacted buildings for source allocation of disaster responses and recovery planning. For example, areas having more major damaged buildings may require more support in the aftermath of disasters.

In this paper, we presented a coupled UNet architecture which uses transformer based difference to perform well on both the damage classification and building segmentation tasks. We applied attention on the difference of the transformer encodings in the feature domain instead of the traditional spatial domain. We hierarchically built the output damage mask by up-sampling and concatenating the low dimension features with higher dimension features. The proposed method yield state of the art results on xBD dataset and LEVIR dataset for damage classification and change detection tasks respectively. We also provided a new baseline for the domain adaptation task on the Ida-BD dataset using xBD dataset. In the future work, we plan to improve the building boundaries using either GAN loss or exponential boundary loss which could help to split different building instances.

References

Bandara, W. G. C., & Patel, V. M. (2022). *A transformer-based siamese network for change detection.* arXiv. Retrieved from `https://arxiv.org/abs/2201.01293`

Benson, V., & Ecker, A. (2020). Assessing out-of-domain generalization for robust building damage detection. *arXiv*, 2011.10328.

Cao, Q. D., & Choe, Y. (2020, September). Building damage annotation on post-hurricane satellite imagery based on convolutional neural networks. *Natural Hazards*, *103*(3), 3357–3376. doi:

Chen, H., Qi, Z., & Shi, Z. (2022). Remote sensing image change detection with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-14.

Cheng, C., Behzadan, A. H., & Noshadravan, A. (2021, February). Deep learning for post-hurricane aerial damage assessment of buildings. *Computer-Aided Civil and Infrastructure Engineering*, mice.12658. Retrieved 2021-05-20, from `https://onlinelibrary.wiley.com/doi/10.1111/mice.12658` doi:

Corbane, C., Carrion, D., Lemoine, G., & Broglia, M. (2011, October). Comparison of Damage Assessment Maps Derived from Very High Spatial Resolution Satellite and Aerial Imagery Produced for the Haiti 2010 Earthquake. *Earthquake Spectra*, *27*(1_suppl1), 199–218. doi:

Fujita, A., Sakurada, K., Imaizumi, T., Ito, R., Hikosaka, S., & Nakamura, R. (2017, May). Damage detection from aerial images via convolutional neural networks. In *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)* (pp. 5–8). doi:

Ghassemi, S., Fiandrotti, A., Francini, G., & Magli, E. (2019). Learning and adapting robust features for satellite image segmentation on heterogeneous data sets. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(9), 6517-6529.

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., … Gaston, M. E. (2019). xBD: A dataset for assessing building damage from satellite imagery. *CoRR*, *abs/1911.09296*. Retrieved from `http://arxiv.org/abs/1911.09296`

Gupta, R., & Shah, M. (2021). Rescuenet: Joint building segmentation and damage assessment from satellite imagery. In *2020 25th international conference on pattern recognition (icpr)* (p. 4405-4411). doi:

Hao, H., Baireddy, S., Bartusiak, E. R., Konz, L., LaTourette, K., Gribbons, M., … Comer, M. L. (2021). An attention-based system for damage assessment using satellite imagery. In *2021 ieee international geoscience and remote sensing symposium igarss* (p. 4396-4399).

Koukouraki, E., Vanneschi, L., & Painho, M. (2022). Few-shot learning for post-earthquake urban damage detection. *Remote Sensing*, *14*(1). Retrieved from `https://www.mdpi.com/2072-4292/14/1/40`

Ku, J., Seo, J., & Jeon, T. (2020). *Dual-hrnet for building localization and damage classification.* Retrieved from `https://github.com/SIAnalytics/dual-hrnet/blob/master/figures/xView2_White_Paper_SI_Analytics.pdf`

Labelbox. (2022). *Labelbox.* Retrieved 2022-05-18, from `https://labelbox.com`

McCarthy, M. J., Jessen, B., Barry, M. J., Figueroa, M., McIntosh, J., Murray, T., … Muller-Karger, F. E. (2020, September). Mapping hurricane damage: A comparative analysis of satellite monitoring methods. *International Journal of Applied Earth Observation and Geoinformation*, *91*, 102134. doi:

Shen, Y., Zhu, S., Yang, T., Chen, C., Pan, D., Chen, J., … Du, Q. (2022). Bdanet: Multiscale convolutional neural network with cross-directional attention for building damage assessment from satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-14.

Spencer, B. F., Hoskere, V., & Narazaki, Y. (2019, April). Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. *Engineering*, *5*(2), 199–222. doi:

Tong, X., Hong, Z., Liu, S., Zhang, X., Xie, H., Li, Z., … Bao, F. (2012, March). 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. doi:

Weber, E., & Kané, H. (2020). Building disaster damage assessment in satellite imagery with multi-temporal fusion. *CoRR*, *abs/2004.05525*. Retrieved from `https://arxiv.org/abs/2004.05525`

Wu, C., Zhang, F., Xia, J., Xu, Y., & Li, G. (2021). Building damage detection using u-net with attention mechanism from pre- and post-disaster remote sensing datasets. *Remote Sensing*, *13*(5). Retrieved from `https://www.mdpi.com/2072-4292/13/5/905`

Zhang, M., Xu, G., Chen, K., Yan, M., & Sun, X. (2019). Triplet-based semantic relation learning for aerial remote sensing image change detection. *IEEE Geoscience and Remote Sensing Letters*, *16*(2), 266-270.