Building Damage Evaluation from Satellite Imagery using Deep Learning

Fei Zhao
Dept. of Computer Science
The University of Alabama at Birmingham
Birmingham, USA
larry5@uab.edu

Chengcui Zhang
Dept. of Computer Science
The University of Alabama at Birmingham
Birmingham, USA
czhang02@uab.edu

***Abstract*—In recent decades, millions of people are killed by natural disasters such as wildfire, landslide, tsunami, and volcanic eruption. The efficiency of post-disaster emergency responses and humanitarian assistance has become crucial in minimizing the expected casualties. This paper focuses on the task of building damage level evaluation, which is a key step for maximizing the deployment efficiency of post-event rescue activities. In this paper, we implement a Mask R-CNN based building damage evaluation model with a practical two-stage training strategy. The motivation of Stage-1 is to train a ResNet 101 backbone in Mask R-CNN as a Building Feature Extractor. In Stage-2, we further build on top the model trained in Stage-1 a deep learning architecture that performs more sophisticated tasks and is able to classify buildings with different damage levels from satellite images. In particular, in order to take advantage of pre-disaster satellite images, we extract the ResNet 101 backbone from the Mask R-CNN trained on pre-disaster images in Stage-1 and utilize it to build a Siamese based semantic segmentation model for classifying the building damage level at the pixel level. The pre- and post-disaster satellite images are simultaneously fed into the proposed Siamese based model during the training and inference process. The output of these two models own the same size as input satellite images. Buildings with different damage levels, i.e., 'no damage', 'minor damage', 'major damage', and 'destroyed', are represented as segments of different damage classes in the output. Comparative experiments are conducted on the xBD satellite imagery dataset and compared with multiple state-of-the-art methods. The experimental results indicate that the proposed Siamese based method is capable to improve the damage evaluation accuracy by 16 times and 80%, compared with a baseline model implemented by xBD team and the Mask-RCNN framework, respectively.**

***Keywords*—Deep Learning, Siamese, Satellite Imagery, Building Damage Evaluation**

I. Introduction and motivation

Natural disasters such as wildfire, landslide, tsunami, and volcanic eruption caused over 1 million of fatalities and 4.4 billion injured cases from 1998 to 2017, despite the great advancement in natural disaster forecasting techniques [1]. With the deadliest natural disasters following one another, the emergency response and humanitarian assistance have become crucial in minimizing the expected casualties. More and more experts in geology, meteorology, and oceanography, etc., participate in the rescue process by analyzing the real-time data and offering their professional suggestions. However, in practice, two obstacles are not supposed to be neglected: during the analysis process, the perspectives of experts often rely on their experience and visual inspections frequently affected by subjectivity and/or subconsciousness. Meanwhile, manual analysis and decision making by human become inefficient and error-prone when the volume of data grows to hundreds and thousands.

With the rapid development of machine learning especially deep learning, it has become feasible that a high-performance computer model extracting high level features from multi-dimensional big data provides experts with the optimal decisions from the objective data automatically. The techniques supporting emergency response and humanitarian assistance have become one of the most popular fields in Data Science community. A number of studies and applications were proposed in the past few years, e.g., supply distribution models, and rescue routing analyzing systems [2–5]. In particular, in recent years, the deep learning methodology allows a computational model to learn high-level features from data in an incremental way. It eliminates the need of domain expertise, which makes it convenient for computer scientists to participate in diversified areas [6]. All of the above contributions that use deep learning to solve practical problems on the emergency response and humanitarian assistance tasks have become a fast growing field in the Computer Science community. An increasing number of models based on deep learning for disaster emergency response tasks have been developed, including the deep learning based rescue scheduling model for disaster management, and the deep learning based helicopter landing site detection model [7–9], etc.

The post-disaster damage level evaluation for residential buildings and commercial buildings is one of the most important components in emergency response tasks, since the damage level of buildings has a critical impact on the expected casualties directly. Moreover, effective damage evaluation on buildings directly impacts the efficiency of the rescue activity deployment and the distribution of rescue supplies. In this paper, the building damage evaluation task is defined as an image segmentation and classification problem. The pixel-wise damage level classification map for buildings in satellite images is the expected output. In fact, there are two commonly adopted contemporary methods that deal with this kind of

978-1-7281-1054-7/20/$31.00 ©2020 IEEE
DOI 10.1109/IRI49571.2020.00020

(a) tornado (b) flooding (c) tsunami

Fig. 1. Some sample pre- and post-disaster satellite images. The images in the top row are satellite images captured before the disasters. The images in the bottom row are obtained after the disaster. In (a) and (c), the red polygons represent the 'destroyed' buildings. In (b), the blue and orange polygons represent the buildings with minor damage and major damage, respectively.

problem using deep learning techniques: U-Net and Mask R-CNN [10, 11]. In general, U-Net is more straight forward and lower in time complexity than Mask R-CNN. However, Mask R-CNN derived from the series of Region Based Convolutional Neural Networks (R-CNN) is a leading architecture of two-stage object detection pipeline [11–14]. The Feature Pyramid Network and Region Proposal Network enable the model to have a good performance on small object detection. For the specific application in this paper, there are four damage levels present on post-disaster buildings, i.e., 'no damage', 'minor damage', 'major damage' and 'destroyed'. Compared with typical object detection applications in Kaggle Challenges, this task can be very challenging since the buildings in satellite images can be very small or even tiny, and the buildings with different damage levels are hard to be differentiated from each other even by visual inspection, especially between minor-damaged and major-damaged buildings. Meanwhile, the variation in building color, shape, and texture, as well as the large visual difference between residential buildings and commercial buildings further increase the difficulty in building detection and (damage level) classification. In our first attempt, we implement a Mask R-CNN framework for classifying the buildings with four different damage levels from post-disaster images. A two-stage training strategy is adopted since the segmentation model trained directly on the post-disaster satellite images is difficult to achieve convergence, despite that the model has been pre-trained on 'COCO' dataset. In Stage-1, a Mask R-CNN model is trained on a simple task to classify buildings from pre-disaster satellite images. The output of the model in Stage-1 has the same size as that of the input pre-disaster image. Each pixel in the output of Stage-1 is assigned a value of either 0 ('background') or 1 ('building'). The training process allows the feature extractor of the model to be sensitive to the building objects including residential houses and commercial buildings. In fact, the Mask R-CNN trained with Two-Stage scheme utilizes the pre-disaster satellite images (Stage-1) to build a building localization model, and then uses the trained model from Stage-1 to further train a damage classification model using post-disaster images only, which has shown much better performance than directly training a Mask R-CNN segmentation model from the post-disaster images alone.

In our second attempt, in order to better and more directly exploit the pair-wise visual contrast between pre- and post-disaster images in damage evaluation, we develop a new two-stage framework based on Siamese neural network, which is proposed for the purpose of signature verification in 1994 [15]. The weight shared sub-networks in Siamese architecture allow the network to extract features from two or more inputs. This architecture enables the network to compare the data/features extracted from the inputs of the respective sub-networks. The motivation of utilizing the Siamese architecture is that

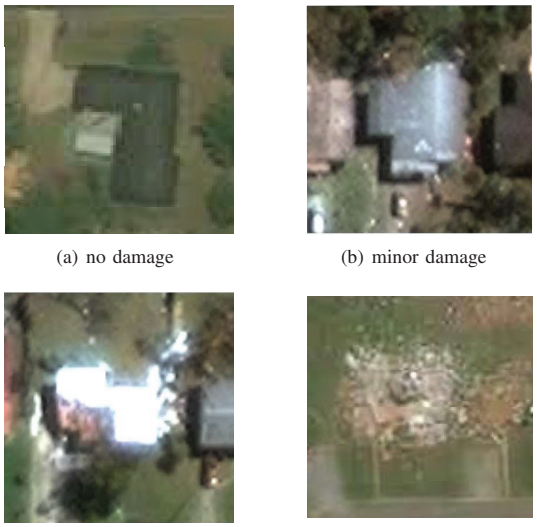
we expect the pre-disaster satellite images to be valuable not only in the training process, but also in the inference process. Therefore, a Siamese based model for pixel-wise classification of buildings with different damage levels is proposed. The model takes advantage of the Mask R-CNN building localization model trained in Stage-1. The ResNet 101 backbone of the Mask R-CNN from Stage-1 is extracted and utilized as the shared-weight feature extractor (encoder part) in the proposed Siamese based model.

The rest of the paper is organized as follows. In Section II, an introduction of the dataset is provided. In Section III, the details of the proposed Siamese based model are presented. In Section IV, the model evaluation metrics is introduced. In Section V, the comprehensive experiments are presented. In Section VI, the results and discussions are delivered. In the last section, the conclusion and the future works are provided.

II. Dataset

In the era of big data, numerous data have been generated over the past few years [16]. For satellite imagery dataset, there are several classic datasets, including AIST Building Change Detection dataset, SpaceNet dataset, and xBD, etc. Most of them are limited to a single disaster type and only contain the post-disaster images for the disaster areas [17]. In this paper, the xBD dataset with 11,034 pairs of pre- and post-disaster images (of size 1024*1024*3) is chosen, which includes 850,736 building annotations. Despite the substantial amount of images and annotations of the dataset, the various disaster types, e.g., flooding, hurricane, and earthquake, etc., and the fine-grained damage levels for buildings, i.e., 'no damage', 'minor damage', 'major damage' and 'destroyed', are the real sparks. Some sample images from the dataset are presented in Figure 1. In this figure, the top row images are some pre-disaster images containing residential houses and commercial buildings. The images in the bottom row are samples of post-disaster images. These six images are sampled from three different types of disasters, i.e., tornado, flooding and tsunami, respectively. From Figures 1(a) and (c), one can easily distinguish 'destroyed' buildings from the other damage levels by simple visual inspection. However, in Figure 1(b), it is pretty hard for people without expertise on building damage evaluation to tell the buildings with minor damage or major damage apart from the others.

The xBD dataset team hired experts to annotate buildings from pre-disaster images, with each polygon representing a building. The ground truth polygons of post-disaster images are the polygons projected from the polygons of pre-disaster images, aligned based on the satellite coordinates. The damage evaluation experts are invited to annotate the polygons in the post-disaster images with the four damage levels. Sample images of the four damage levels are shown in Figure 2. In the annotation process for post-disaster images, the evaluation standard can be concluded as follows: buildings with no sign of water, structural damage or burn marks are labeled as no damage. The buildings labeled as destroyed are completely collapsed or covered with water or mud. The minor damage

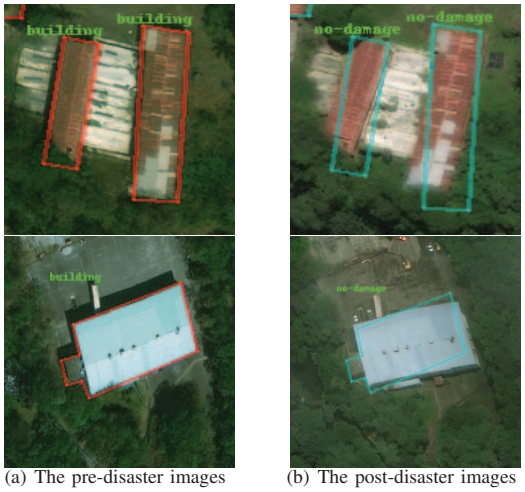


(a) no damage (b) minor damage

(c) major damage (d) destroyed

Fig. 2. The different damage levels

class is assigned based on whether the building is partially burnt, water surrounding, roof elements missing, or contains visible cracks. And the major damage focuses on the partial wall or roof collapse, or buildings surrounded by water. While there exists a huge difference between no damage buildings and destroyed buildings, it is often hard to tell the minor damaged buildings apart from the major damaged buildings by visually inspecting the satellite images [17].



(a) The pre-disaster images (b) The post-disaster images

Fig. 3. The misalignment issues

Since the building polygons of post-disaster images are obtained from the projection of polygons in their pre-disaster counterparts, there is an issue of misalignment between the polygons in the pre-event and post-event disaster samples due to problems such as shifting or tilting. Some samples are

shown in Figure 3. In this figure, the Red Polygons are the localization ground truth of the buildings in the pre-disaster images. The Blue Polygons and labels of damage levels are the 'ground truth' of the buildings in the post-disaster images. Because of the misalignment issues, training the model only on the post-disaster is not ideal and according to our experiments has led to poor performance. Some samples of pre- and post-disaster ground truth images are shown in Figure 4, in which different colored segments represent buildings with different levels of damage.

(a) The pre-disaster image (b) The post-disaster image

Fig. 4. The ground truth of pre- and post-disaster images. In (a), the polygons represent the building locations. In (b), the Green, Blue, Orange and Red colors represent the four different damage levels: no damage, minor damage, major damage and destroyed, respectively.

In addition to the misalignment issues, the imbalance in the dataset is not supposed to be neglected. The major issue is that the buildings with no damage are the majority class in the original dataset. The size of this majority class is at least one order of magnitude larger than any other class. In order to deal with this issue, we build a new dataset by steps as follows: firstly, from the original dataset, we identify 6,480 original images (no duplicate item) containing at least one minor, major or destroyed building to form the new basic dataset. Secondly, the basic dataset is split into 9 : 1 to form the basic training dataset and the validation dataset, respectively. Thirdly, from the basic training dataset, we find the set of images containing at least one minor or major damaged building (minor-major group), and the set dominated by major damaged or destroyed buildings (major-destroyed group). We then add one copy (with variations, e.g., rotate, shift, re-scale) of each image in the minor-major group into the training dataset, and also add one copy (with variations) of each image in the major-destroyed group into the dataset. For example, if an image contains minor, major, and destroyed buildings, and is 'destroyed' dominated, there will be three (varied) copies of this image including the original one in the training dataset. However, if it is minor-damage dominated, then only two copies including the original image will be in the augmented dataset. Since the results of model will be compared with the baseline model on the untouched data, we utilize the additional 1,866 images provided by the xBD team as the test dataset [17]. The size of the dataset is shown in Table I.

TABLE I
THE AUGMENTED DATASET FOR MODEL TRAINING AND TESTING.

	Initial Size	After Up-sampling
Training Dataset	5,840	12,030
Validation Dataset	640	640
Test Dataset	1,866	1,866

III. THE PROPOSED MODEL

The motivation of implementing Siamese Architecture for damage evaluation is that utilizing both pre- and post-disaster images can benefit the training and inference process. The difference between pre- and post-disaster images is expected to be learned automatically by the training process and helps the model to better predict the damage level of buildings in the inference process.

As mentioned above, a Mask R-CNN framework is trained with a two-stage strategy. The purpose of Stage-1 is to obtain a high level feature extractor (ResNet 101) from a trained Mask R-CNN that is sensitive to the localization of buildings. Here, the proposed Siamese based model takes advantage of this feature extractor. The ResNet 101 backbone from the Mask R-CNN, which has completed stage-1 training process, is set up as the weight-shared high level feature extractor in the Siamese based model. It allows the Siamese based model to obtain the valuable feature data from the pre- and post-disaster images at the very beginnings of the training process for classifying building damage levels. While the proposed model is named Siamese based framework, there are two major differences between the proposed model and the traditional Siamese model. Firstly, the proposed model is an end to end semantic segmentation framework without any fully connected layers, while the traditional Siamese Architecture Network is designed for image classification problems trailed by fully connected layers. This means that there is no expectation on fixed-size inputs for the proposed Siamese based model. Second, the proposed model does not consist of any hard coded calculations, while the traditional Siamese Architecture contains the L1 distance or L2 distance layer for measuring the difference between the two inputs. In our model, we expect it to be able to learn complex relationship between the features extracted from the pre- and post-disaster image pipeline. The structure of the proposed model is presented in Figure 5.

As Figure 5 shows, the proposed model consists of three main parts: the first part, the ResNet 101 extracted from Stage-1 trained Mask R-CNN is implemented as the encoder part of the model. It provides the model with the extracted high level features from pre- and post-disaster images. The second part, the decoder part of the model consists of repeated application of two 3x3 convolutional layers with the rectified linear unit and a up-sampling layer [10]. The third part, the skip connections are implemented for boosting the recovery of the full spatial resolution. Since detecting small buildings in large-scale satellite images is difficult, the loss function combines the cross entropy loss, the dice loss and focal loss.

The cross entropy loss can be formulated as Equation (1),

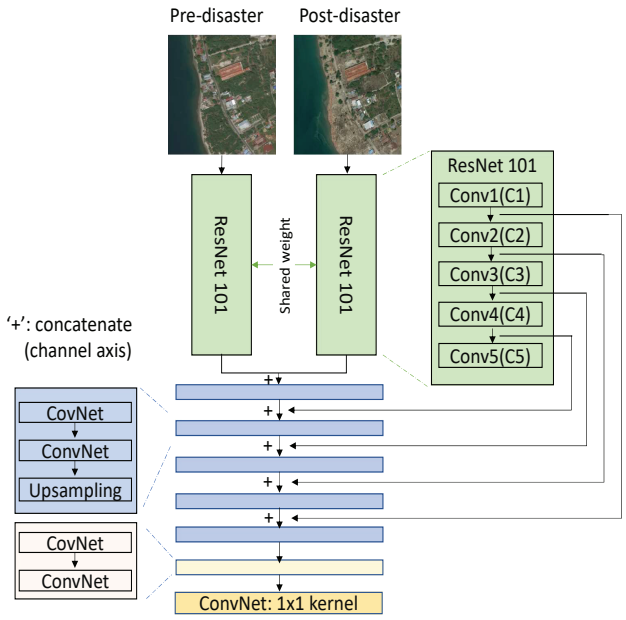


Fig. 5. The proposed Siamese based model

where n, m, $y_{i,j}$, and $p_{i,j}$ represent the number of samples, the number of classes, the ground truth label, and the predicted probability for the sample i being class j, respectively.

$$L_{ce} = -\sum_{i}^{n}\sum_{j}^{m} y_{i,j} log(p_{i,j}) \tag{1}$$

The focal loss is one of the most popular losses for image segmentation problems. It can be formulated as Equation (2), where $\gamma \geq 0$, and α is a weight hyper-parameter. When an example is misclassified and $p_{i,j}$ is small, the loss is unaffected. However, when an example is well-classified, and $p_{i,j}$ is close to 1, the loss for this well-classified example is down-weighted [18].

$$L_{focal} = -\sum_{i}^{n}\sum_{j}^{m} \alpha_j (1-p_{i,j})^{\gamma_j} y_{i,j} log(p_{i,j}) \tag{2}$$

Dice loss is a region-based loss function. It aims to minimize the mismatch regions and maximize the overlap regions between ground truth and predicted segmentation [19]. The advantage of dice loss is that it directly optimizes the dice coefficient (F1 score) which is used as the evaluation metric in this paper. Therefore, dice loss is included in our loss function, and defined in Equation (3), where ϵ is a small positive infinitesimal quantity.

$$L_{dice} = \sum_{j}^{m}(1 - \frac{2\sum_{i}^{n} p_{i,j} y_{i,j} + \epsilon}{\sum_{i}^{n}(p_{i,j} + y_{i,j}) + \epsilon}) \tag{3}$$

The definition of the total loss for the proposed Siamese based model is shown in Equation (4):

$$Loss_{total} = L_{ce} + L_{focal} + L_{dice} \tag{4}$$

IV. Metrics

F1 score is chosen as the basic component of the evaluation metrics, since it incorporates both precision and recall. The definition is as follows:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5}$$

In Equation (5), the calculation of $Precision$ and $Recall$ involves $True\ Positive$ (TP), $False\ Positive$ (FP), and $False\ Negative$ (FN) as defined below. Figure 7 shows an example of how to calculate TP, FP, and FN.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{6}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{7}$$

The four damage levels, i.e., no damage, minor damage, major damage, and destroyed, are each assigned a value 1, 2, 3, and 4, respectively. We follow the $F1_{damage}$ definition from [17] in this study as defined in Equation (8), in which $F1_i$ where $i \in [1, 4]$ represents the F1 score of each damage level. The finial F1 score is the harmonic mean of the F1 scores of all four classes. The ϵ is a small positive infinitesimal quantity.

$$F1_{damage} = \frac{4}{\sum_{i=1}^{4}(F1_i + \epsilon)^{-1}} \tag{8}$$

V. Experiments

The comparative experiment is conducted on the dataset described in Section II. In the experiment, there are three models in comparison: the baseline model [17], the 2-stage Mask R-CNN model as introduced in Section I (first attempt), and the proposed Siamese based model (second attempt). The damage classification baseline model is provided by the xBD team. The architecture of the baseline model is presented in Figure 6 [17]. In [17], the buildings in post-disaster images are each cropped out and scaled to a small image of 128*128*3, based on their ground truth polygon coordinates. The cropped building images are then fed into the two branches of the baseline model (Figure 6). The concatenated high level features are passed to the dense layers for classifying the images. In [17], the authors not only provide the implementation of the baseline model, but also report the classification results on the test dataset.

The Mask R-CNN framework with ResNet 101 backbone is implemented as one of the comparative models, following the two-stage training strategy introduced in Section I. In Stage-1, the model is initialized with weights pre-trained on the MS COCO dataset, and is trained on the pre-disaster images

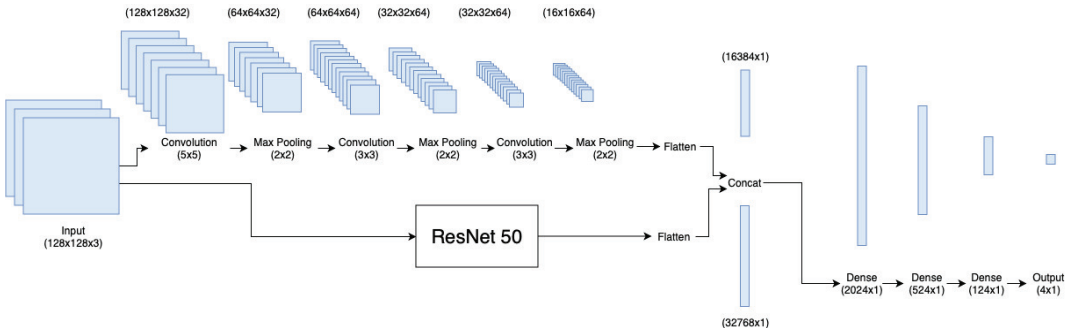


Fig. 6. The architecture of the baseline model

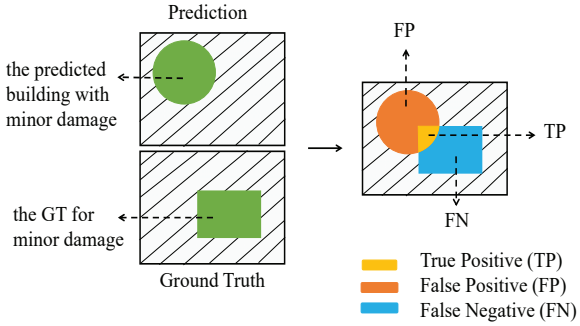


Fig. 7. The calculation process for True Positive (TP), False Positive (FP), and False Negative (FN). This figure shows an example for the minor damage. For the other damage levels, to obtain the TP, FP, and FN follows the same pipeline.

only. In order to obtain the best performing model and also for fair comparison, the model is tuned with different hyper-parameter values. We performed several hyper-parameter tuning operations including varying the size of the anchors for the Region Proposal Network (RPN), initializing with different learning rates and learning momentum, and modifying the non-max suppression threshold to filter RPN proposals so that more proposals can be generated during training process, etc. Moreover, the data pre-processing is applied to normalize the image data. In Stage-2, the model trained in Stage-1 is further trained on post-disaster images to classify buildings into four different damage levels. We also tuned the learning rate, learning momentum, and batch size, etc. Nevertheless, since some hyper-parameters for inference process are critical to the final performance of the model, we further tuned these to enhance the performance of the Mask R-CNN for the specific application, including minimizing the threshold for detection confidence score, increasing the maximum threshold of the number of detection instances, and decreasing the non-maximum suppression threshold for detection process, etc. By tuning the hyper-parameters of the model, the F1 score of Mask R-CNN is improved by at least 4 times compared with the initial version.

The proposed Siamese Based Model also takes advantage of the Stage-1 well-trained Mask R-CNN model. The ResNet 101 backbone in the Mask R-CNN is extracted and incorporated into the Siamese based model as the high level feature extractor. The model is trained on the dataset mentioned in Section II. Data pre-processing (normalization) is implemented the same as before. For both the Siamese based model and the Mask R-CNN model in Stage-2, the data augmentation is implemented, including the up-down flip, the lift-right flip, the rotation with 90 degree, 180 degree and 270 degree, etc. Both the Siamese based model and the Mask R-CNN model in Stage-2 are trained on a single NVIDIA Tesla P100 16GB GPU with 50 epochs.

VI. Results and Discussions

The F1 scores of the baseline model, the Mask R-CNN framework, and the proposed Siamese based model are shown in Table II and Figure 8. The harmonic mean of F1 scores of all the four damage classes obtained by the proposed Siamese based model is 0.5732, which is 16 times and 80% better than that of the baseline model and the Mask R-CNN framework, respectively. Moreover, the F1 score for each individual class obtained by the proposed Siamese based model is higher than that of the other two models.

TABLE II
The F1 score of each model

F1 Score	Baseline	Mask R-CNN	Siamese Based Model
no damage	0.6631	0.5728	**0.7140**
minor damage	0.1435	0.1648	**0.3955**
major damage	0.0094	0.3642	**0.6037**
destroyed	0.4657	0.4977	**0.7181**
Mean	0.3204	0.3999	**0.6078**
Harmonic Mean	0.0342	0.3183	**0.5732**

Based on our initial visual inspections, it is expected that it would be much easier to classify no-damage and destroyed buildings, which is proven through the experimental results. However, surprisingly, the classification accuracy of major-damaged buildings is not far behind.

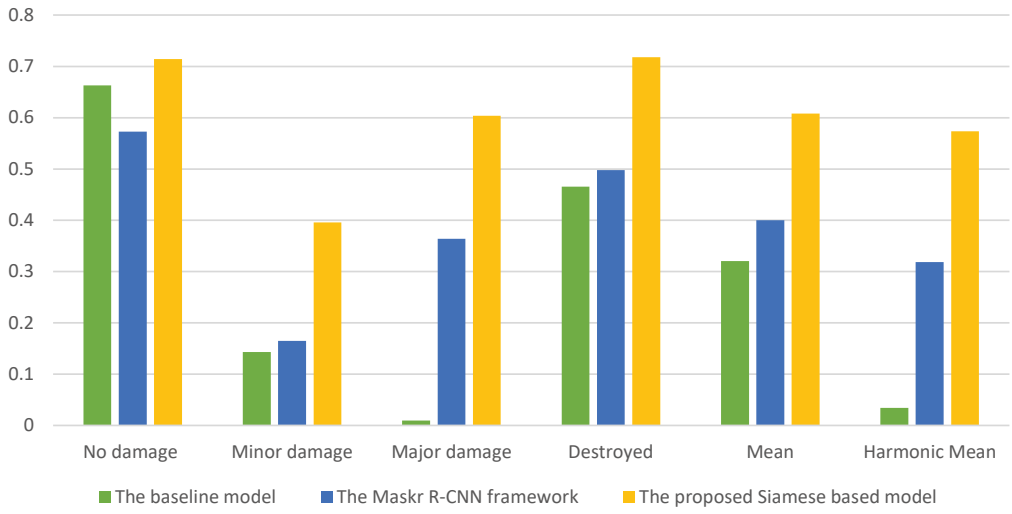


Fig. 8. The F1 score of each model

Several sample and potentially interesting model outputs are shown in Figure 9. In Figure 9(a), the Siamese based model produces a near-perfect prediction for the buildings with major damage. In Figure 9(b), the buildings with no damage and minor damage are predicted fairly accurately by the Siamese based model. However, in Figure 9(c), while the model predicts reasonable results for the buildings in destroyed category, in the upper-right corner there are false positives for no-damage class (green) and some false negatives for major-damage class (blue). The prediction results shown in Figure 9(d) are much worse than the other three cases.

These cases indicate that the proposed model may perform better when there is one single dominating class in an image. However, images containing crowded buildings with various damage levels and no dominating class can be hard for the proposed model to work on. This might be caused by the receptive field limitation in the proposed Siamese based model. The feature obtained from the last layer of the pre-disaster encoder pipeline is the high level feature. The large receptive field makes it difficult to recover the low-level details in the crowded area. It seems possible that concatenating more features extracted from different layers of the pre-disaster encoder pipeline with the decoder part of the Siamese based model might be able to improve the performance in dense regions.

VII. CONCLUSIONS AND FUTURE WORKS

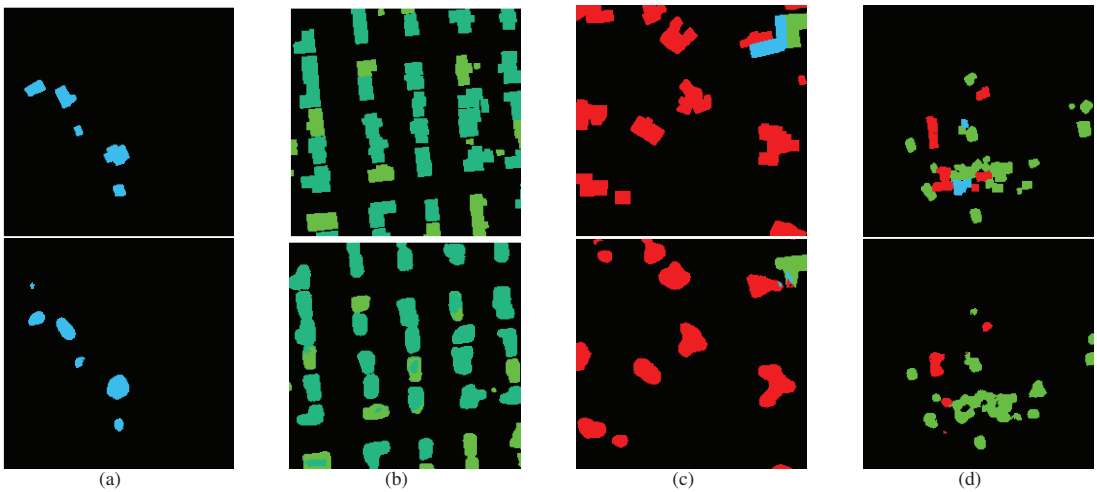
In this paper we propose a two-stage, Siamese based deep learning framework for evaluating building damage levels from satellite images. In Stage-1, a Mask R-CNN model is trained to localize building objects from pre-disaster buildings and more importantly to serve as a feature extractor on top of which a modified Siamese model is further trained in Stage-2 for building damage level classification. The proposed Siamese based model is fed with both pre- and post-disaster images and is capable to produce more accurate classification maps than the baseline model [17] and a two-staged Mask R-CNN model. The experimental results indicate that the Siamese based method improves the building damage classification accuracy by 16 times and 80%, compared to the baseline model and the two-staged Mask R-CNN model, respectively.

In the future, we see at least three ways to further improve the classification accuracy for building damage evaluation: first, we plan to modify the proposed model by concatenating multi-level features extracted from the pre-disaster encoder pipeline with the decoder part. With this modification, we expect to be able to test and tell thoroughly whether injecting more features from different levels of the encoder into the decoder can improve the classification accuracy in this particular application. Second, we plan to incorporate an attention mechanism into the proposed framework which is expected to automatically learn to keep focus on the most important/interesting regions of the image. We expect this pipeline will alleviate the problems with small or tiny object detection tasks. Last but not the least, based on our experience and observations obtained from this project, we plan to tackle this problem from a completely new angle by modeling it as a deep regression problem which considers the relative distance between and among the four damage levels.

REFERENCES

[1] P. Wallemacq, *Economic losses, poverty & disasters: 1998-2017*. Centre for Research on the Epidemiology of Disasters, CRED, 2018.

[2] Y. X. Chen, P. R. Tadikamalla, J. Shang, and Y. Song, "Supply allocation: bi-level programming and differential evolution algorithm for natural disaster relief," *Cluster Computing*, vol. 23, no. 1, pp. 203–217, 2020.



(a) (b) (c) (d)

Fig. 9. Some interesting cases. In this figure, the colors Green, Dark Green, Blue, and Red represent: no damage, minor damage, major damage and destroyed, respectively. The top row images are ground truth. The bottom row images are damage predictions from the proposed Siamese based model.

[3] V. Hristidis, S.-C. Chen, T. Li, S. Luis, and Y. Deng, "Survey of data management and analysis in disaster situations," *Journal of Systems and Software*, vol. 83, no. 10, pp. 1701–1714, 2010.

[4] B. Liu, J. B. Sheu, X. Zhao, Y. Chen, and W. Zhang, "Decision making on post-disaster rescue routing problems from the rescue efficiency perspective," *European Journal of Operational Research*, pp. 321–335, 2020.

[5] S.-C. Chen, M.-L. Shyu, C. Zhang, W. Z. Tang, and K. Zhang, "Damage pattern mining in hurricane image databases," in *Proceedings of the Fifth IEEE Workshop on Mobile Computing Systems and Applications*, pp. 227–234, 2003.

[6] M. Krishnamoorthy, S. Suresh, and S. Alagappan, "Deep learning techniques and optimization strategies in big data analytics: automated transfer learning of convolutional neural networks using enas algorithm," in *Deep Learning Techniques and Optimization Strategies in Big Data Analytics*, pp. 142–153, 2020.

[7] D. T. Nguyen, S. Joty, M. Imran, H. Sajjad, and P. Mitra, "Applications of online deep learning for crisis response using social media information," *arXiv preprint arXiv:1610.01030*, 2016.

[8] M. Y. Kabir and S. Madria, "A deep learning approach for tweet classification and rescue scheduling for effective disaster management," in *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 269–278, 2019.

[9] V. Antoniou and C. Potsiou, "A deep learning method to accelerate the disaster response process," *Remote Sensing*, vol. 12, no. 3, p. 544, 2020.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, 2015.

[11] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, 2014.

[13] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, pp. 91–99, 2015.

[15] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a" siamese" time delay neural network," in *Advances in neural information processing systems*, pp. 737–744, 1994.

[16] B. Data, "for better or worse: 90% of world's data generated over last two years", sintef, 22 may 2013," *Science Daily, http://www. sciencedaily. com/releases/2013/05/130522085217. htm*.

[17] R. Gupta, R. Hosfelt, S. Sajeev, N. Patel, B. Goodman, J. Doshi, E. Heim, H. Choset, and M. Gaston, "xbd: A dataset for assessing building damage from satellite imagery," *arXiv preprint arXiv:1911.09296*, 2019.

[18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[19] J. Ma, "Segmentation loss odyssey," *arXiv preprint arXiv:2005.13449*, 2020.