

RescueNet: Joint Building Segmentation and Damage Assessment from Satellite Imagery

Rohit Gupta and Mubarak Shah
Center for Research in Computer Vision

University of Central Florida

Orlando, Florida 32816

Email: rohitg@knights.ucf.edu, shah@crcv.ucf.edu

Abstract—Accurate and fine-grained information about the extent of damage to buildings is essential for directing Humanitarian Aid and Disaster Response (HADR) operations in the immediate aftermath of any natural calamity. In recent years, satellite and UAV (drone) imagery has been used for this purpose, sometimes aided by computer vision algorithms. Existing Computer Vision approaches for building damage assessment typically rely on a two stage approach, consisting of building detection using an object detection model, followed by damage assessment through classification of the detected building tiles. These multi-stage methods are not end-to-end trainable, and suffer from poor overall results. We propose RescueNet, a unified model that can simultaneously segment buildings and assess the damage levels to individual buildings and can be trained end-to-end. In order to model the composite nature of this problem, we propose a novel localization aware loss function, which consists of a Binary Cross Entropy loss for building segmentation, and a foreground only selective Categorical Cross-Entropy loss for damage classification, and show significant improvement over the widely used Cross-Entropy loss. RescueNet is tested on the large scale and diverse xBD dataset and achieves significantly better building segmentation and damage classification performance than previous methods and achieves generalization across varied geographical regions and disaster types.

I. INTRODUCTION

In the wake of natural disasters, resources available to first responders are scarce, and efficient planning and allocation of aid and rescue efforts can help save thousands of lives. Traditionally, response planning has been based on reports and estimates based on ground based assessments. Ground based assessments are risky and potentially impossible to obtain, hence, more recently aerial and satellite imagery has been used for these assessments [2]. While analysis of satellite and aerial imagery by experts is useful for rapid response operations, it still results in time lags that could otherwise be spent on rescue operations, as even large teams can take weeks to completely map out disaster affected areas [3]. Automated methods for analyzing aerial and satellite imagery have been developed, including those relying on handcrafted rules for identifying damaged buildings from LiDAR point clouds [4], segmenting the perimeter of forest fires using deep learning [5], detecting flooded regions [6], detecting collapsed, and damaged buildings using Convolutional Neural Networks [7] [8] and detecting damaged buildings using object detectors [9].

In this paper, we focus on assessing damage levels for buildings, which is relevant to all types of natural disasters,

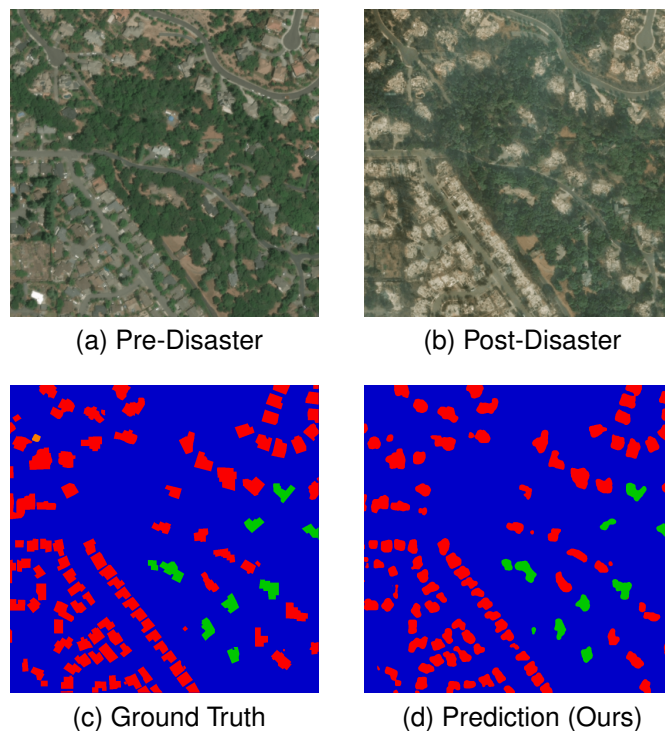


Fig. 1. Samples from the xBD dataset [1] for building damage assessment. (a) and (b) respectively show Pre and Post Disaster images. (c) and (d) shows ground truth and prediction from proposed method. Red labels represent completely damaged buildings, and Green labels represent undamaged buildings. Intermediate damage levels are represented by pink (major damage) and orange (minor damage).

and can have a significant impact on the efficacy of search and rescue operations in their aftermath. State of the art methods for detecting damaged buildings [10] [1] rely on a two stage pipeline, where buildings are detected in the pre-disaster imagery in the first stage, and then detected building are classified into different damage levels by comparing pre and post disaster imagery in the second stage. These multi-stage methods are not end-to-end trainable, and suffer from poor overall results. In contrast, we propose RescueNet, which is an end-to-end trainable, unified model to segment buildings and classify their damage levels in one go. We employ a pixel-level segmentation based approach, and use multi-scale,

temporal (capturing before and after damage) features to detect damage in the segmented buildings. To model the hierarchical and composite nature of this problem, we also design a novel loss function which provides a significant improvement over the standard cross-entropy loss. To benchmark our method, we compare against an existing baseline on the public and large scale xBD dataset [1], which contains images from a variety of disaster events such as earthquakes, flooding, hurricanes and forest fires from 19 locations across the world. This diversity allows for training and testing the model on a wide variety of damage types. RescueNet is able to achieve significantly improved performance over the baseline, especially on damage classification. Sample results are shown in Figure 1 and more results are shown in Figure 3.

The organization of the rest of the paper as follows. Section II provides a brief overview of the existing literature on the problems of Building Footprint Segmentation, Building Damage Assessment and Change Detection in remote sensing using satellite imagery. Section III presents the architecture of our model, and our training approach. Section IV provides a brief summary of the characteristics of the xBD dataset. Section V contains an analysis of the results obtained by our method. Finally, our conclusions are provided in Section VI.

II. RELATED WORK

A. Building Footprint Segmentation

Most prior works on training deep building footprint detection models use the DeepGlobe [11] or SpaceNet [12] datasets, while others like BingHuts [13] and the ISPRS 2D Semantic Labelling (Vaihingen and Potsdam) [14] dataset also exist. Whereas, SpaceNet and DeepGlobe are more diverse and large scale, hence better suited for training and evaluating deep learning methods. State of the art methods for Building footprint segmentation on these datasets typically rely on models like U-Net [15], originally developed for semantic instance segmentation problems. TernaNetv2 [16] is a modified U-Net to operate on multi-spectral imagery, and to predict instance boundaries along with pixelwise semantic segmentation. Hamaguchi *et al.* [17] utilize ensembles of U-Nets, each of which is trained to detect buildings of a specific size range. A variant of LinkNet (a U-Net with residual connections in the encoder), [18] with SE-ResNext backbone, and a composite loss function consisting of the binary cross-entropy loss, combined with the Lovász hinge loss and mean squared watershed energy loss is used in Golovanov *et al.* [19].

B. Change Detection

There has been an extensive amount of research on change detection by remote sensing researchers prior to the advent of deep learning. Typically, these methods applied pixel difference based models to long term data sets of geometrically and radiometrically corrected satellite images. Radke *et al.* [20] presents a comprehensive overview of this class of change detection methods. More recently, Deep CNN based models have been used to classify the difference images for change detection in [21]. Using Siamese networks for pixel wise

change detection has been studied in [22]. Daudt *et al.* [23] use an iterative refinement and training procedure to learn a change detection model from noisy data. Recurrent Neural Networks have also been used for Change Detection in Multi-Temporal data [24].

C. Building Damage Assessment

While the literature on building footprint segmentation and change detection is very extensive, the problem of building damage assessment from satellite imagery has only received limited attention. Xu *et al.* [10] investigate a two stage architecture for detecting damaged buildings. A Faster R-CNN [25] model is trained to detect building tiles, and a change detection network is trained as a binary classifier on pre and post disaster building tile pairs. Various simple CNN architectures are tried for the change detection network, and the architecture using feature differences is found to be the best. Gupta *et al.* [1] provide baseline results for the xBD dataset. They adopt a U-Net based model originally designed to detect building footprints in SpaceNet for their first stage. Additionally, for change detection, they use a two-stream classification approach.

A crucial limitation of these methods is the use of separate stages for detecting buildings and damage classification. This necessitates the use of greedy stage wise training, which makes it impossible for the model to benefit from multi-task supervision. RescueNet explores the possibilities offered by a joint segmentation and damage assessment framework and aims to close this gap in the literature.

III. RESCUNET MODEL

RescueNet is a unified model for building segmentation and damage classification. A schematic block diagram showing components of RescueNet's architecture grouped into logical modules is provided in Figure 2. The design of each of these modules and explanations for the design choices made therein are provided further on in this section. The design of the feature extractor and segmentation head draw on the semantic segmentation literature, specifically DeepLabv3 [26] and DeepLabv3+ [27], adapting them to the task of segmenting buildings, and augmenting it with the capability of change detection across multi-temporal images.

A. Extracting Image Features

The goal of the first stage of the network is to extract multi-scale image features at a high spatial resolution. In order to preserve higher spatial resolution in the features RescueNet adopts a backbone architecture consisting of a ResNet50 [28], with dilated convolutions replacing strided convolutions beginning from the second ResNet Blocks [29]. This results in features having 1/8th the spatial resolution of the input image along both height and width. To obtain multi-scale features Átrous Spatial Pyramid Pooling (ASPP) module is used on top of the backbone CNN features. The ASPP module uses image pooling and dilated convolutions of kernel size 3 and dilation sizes of 12, 24 and 36 to obtain features at 4 different scales.

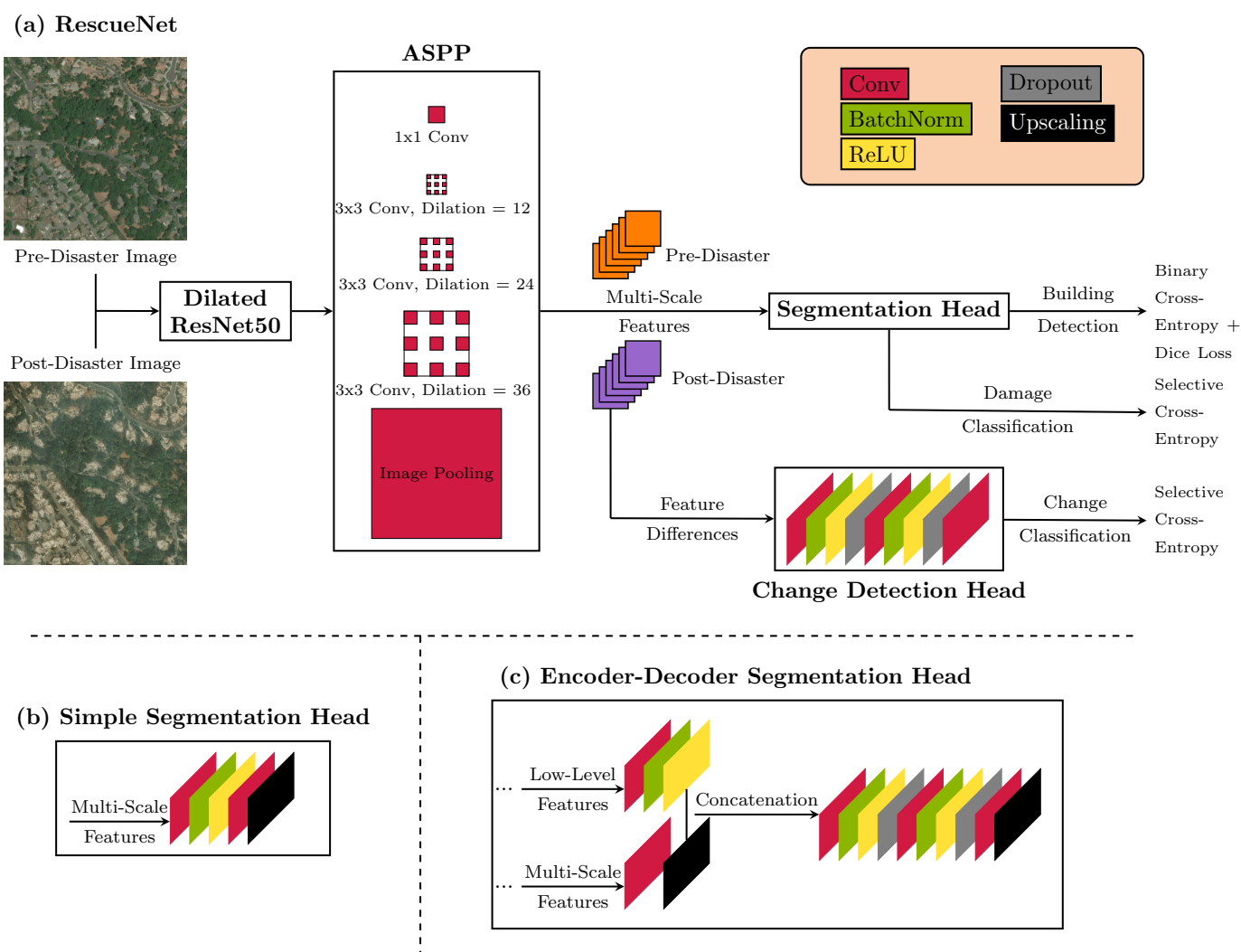


Fig. 2. (a) **RescueNet Model**. The backbone of the network is a Dilated ResNet, which produces features at 1/8th input image resolution. Atrous Spatial Pyramid Pooling module uses dilated convolutions of strides 12, 24 and 36, along with image pooling to generate multi-scale features. Segmentation head can have either the simple convolutional design or follow the encoder-decoder style. The segmentation head makes predictions independently for pre- and post-disaster images. Change Detection head applies Convolutional layers with batch normalization and ReLU activations followed by an output 1x1 Convolution to the difference of post- and pre-disaster features. The xBD dataset has 4 different damage classes ($C=4$), while building segmentation is a binary problem. (b) **the simple segmentation head**. (c) **the encoder-decoder style segmentation head**.

B. Segmentation Head

We experiment with two possible architectures for the segmentation head. The simple architecture consists of a convolution block, whereas the Encoder-Decoder architecture additionally has skip connections from the backbone network's lower level features. In both cases these layers are followed by upsampling to make segmentation predictions at full image resolution. The Segmentation head classifies the pixels of the pre- and post-disaster images independently, hence for damage classification, it only processes post-disaster imagery.

C. Change Detection Head

In order to utilize the temporal (post and pre-damage) features available from the input, we design an additional change detection head. The change detection head is a simple

Convolution block with 3 layers, and utilizes as its input the difference of multi-scale features from the pre- and post disaster images. We utilize the difference of features as the amount of change is expected to be associated with the distance in the semantic feature space. The change detection head classifies each pixel into the four damage categories, and the predictions made for background pixels are ignored while training and during inference.

D. Loss Functions

Choosing a good loss function is very important and hence we experiment with a variety of relevant loss functions. To begin with we use the simple Cross-Entropy Loss on top of the segmentation output to train the network. However, the cross-entropy loss is not well suited for this problem because the network needs to classify pixels at 2 different levels: building

localization and damage classification. With this insight in mind, we developed a localization aware loss function, which accounts for the hierarchical nature of the problem. The localization aware loss consists of a Binary Cross Entropy loss for building segmentation, and a foreground only selective Categorical Cross-Entropy loss for damage classification. Since this loss matches the structure of the problem better, we expect it to perform significantly better than plain Cross-Entropy. Formally, the localization aware loss function is defined in Equation 1. In the equation, y_{il} is the ground-truth binary localization label indicating whether the current pixel belongs to a building or not, y_{ik} is the damage classification label for class k , and \hat{y}_{il} and \hat{y}_{ik} are the predicted values for them. The damage classification loss is summed over the set of all damage classes C .

$$L(y_i, \hat{y}_i) = \begin{cases} -\log(\hat{y}_{il}) + \sum_{k \in C} -\log(\hat{y}_{ik}) & \text{if } y_{il} = 1 \\ -\log(1 - \hat{y}_{il}) & \text{if } y_{il} = 0 \end{cases} \quad (1)$$

Additionally, in order to predict crisper building boundaries we also employ with using the Dice Loss [30] [31] for building segmentation. The Dice Loss is based on the Sørensen–Dice coefficient:

$$DiceLoss(y_{il}, \hat{y}_{il}) = \frac{2|y_{il} \cap \hat{y}_{il}|}{|y_{il}| + |\hat{y}_{il}|}. \quad (2)$$

Since y_{il} , the ground truth building mask label is binary, the intersection term in the numerator can be computed as the product of ground-truth and prediction, and the denominator is the sum of the absolute values, which means the loss is piece-wise differentiable and hence trainable.

IV. DATASET

We use xBD [1], which is the only large-scale public dataset for building segmentation and damage assessment to benchmark our method. xBD has been collected for the purpose of aiding research that results in technology to assist Humanitarian Aid and Disaster Relief Efforts. It consists of high quality building segmentation and damage assessment annotations for high-resolution satellite imagery, collected before and after 19 disaster events (such as floods, volcanic eruptions, earthquakes, and hurricanes) spread across the world. The dataset uses polygons to represent building segments and a novel 4 point damage scale. 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. 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. Sizes of each dataset split can be found in Table I. For reproducibility,

the list of data samples in the validation split will be released with the supplementary material.

TABLE I
SIZE OF XBD SPLITS

Data Split	Image Pairs
Tier1 Train	2,495
Tier3 Train	6,369
Tier1 Validation	304

V. RESULTS

In this section we present the results of benchmarking our model on the xBD dataset. We first present results from our model for some samples from the validation set, then define the XView Challenge metric which is used as the primary quantitative metric, then present ablation studies in order to study the impact of various components of our method, followed by a comparison of our results with the baseline.

A. Qualitative Results

In Figure 3 we present some qualitative results on a small but diverse sample of the dataset. The selected images are from areas with significant diversity in density of buildings and type of damage. As can be noted from these results, RescueNet appears to be significantly better than the baseline model at predicting the damage levels. The baseline model makes quite a few false positive and false negative errors that are eliminated by our model. Significant variation can be observed in the pattern of damage caused by different kinds of disasters, with flooding affecting neighbouring buildings to widely varying extents, whereas disasters like Tsunamis which strike with concentrated force destroy a contiguous region. The qualitative results also suggest two sources for the remaining error of our model: first, extremely large buildings not being recognized as one unit by the damage classification head, and secondly, the building segmentation head not being able to separately resolve neighbouring buildings that are extremely close to each other.

B. Metrics

We report results using the XView2 Challenge metric¹, a weighted average of the building segmentation F1 score and the harmonic mean of class wise damage classification F1 scores:

$$Score = 0.3 * F1_{loc} + 0.7 * \frac{n}{\frac{1}{F1_{cls1}} + \dots + \frac{1}{F1_{cls n}}} \quad (3)$$

Here, $F1_{loc}$ is the F1-Score for building segmentation and $F1_{cls1} \dots F1_{cls n}$ are the damage classification F1-scores for each damage class. This metric is very challenging because it heavily penalizes overfitting to overrepresented classes and the xBD dataset is heavily skewed.

¹<https://xview2.org/challenge>

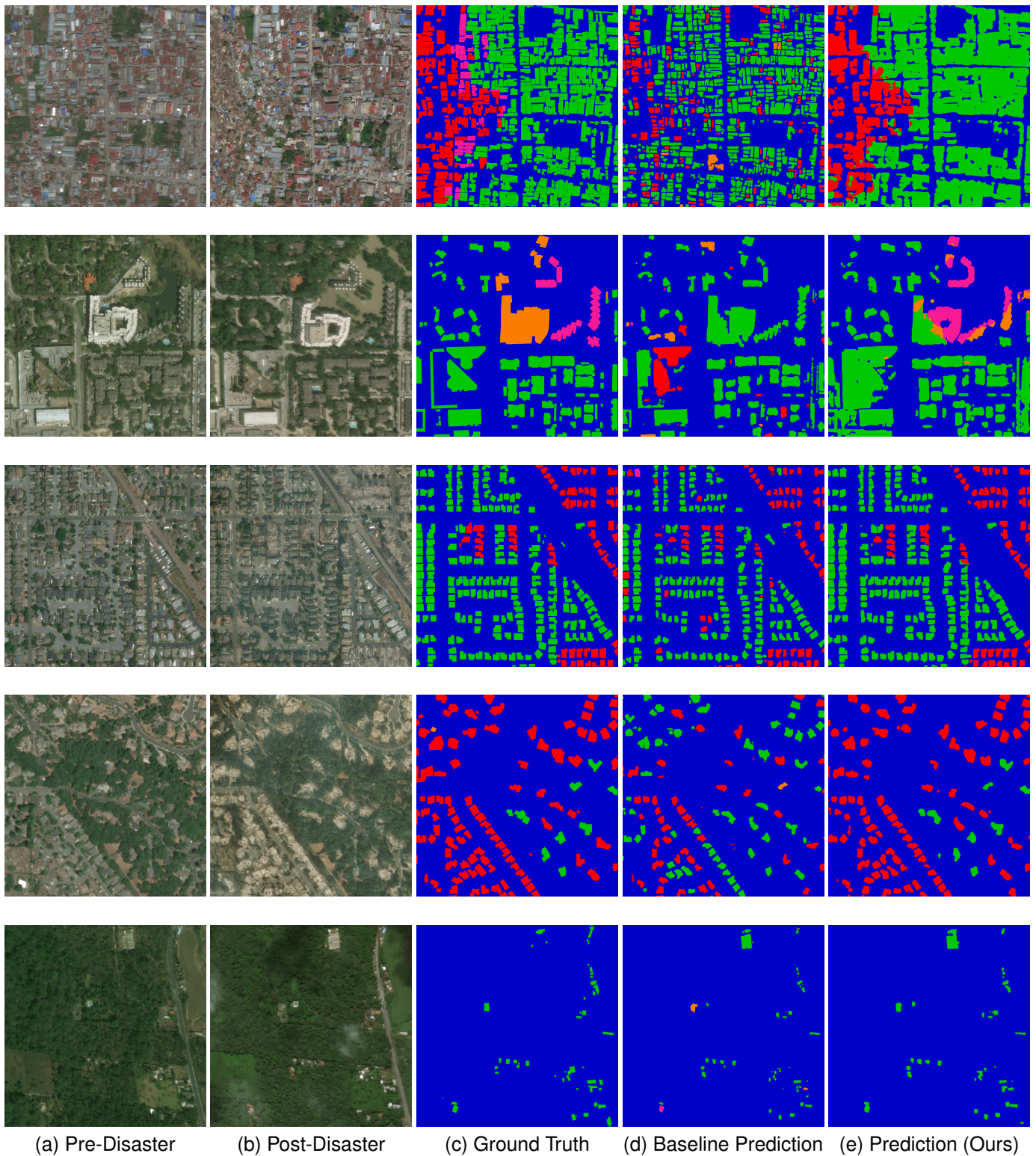


Fig. 3. Qualitative Results for a diverse set of examples from the Validation set. **From Left to Right in each row:** Pre-Disaster Image, Post-Disaster Image, Ground Truth Labels, Predicted labels (Baseline Model), Predicted labels (RescueNet). **From Top to Bottom:** A very high density urban area severely hit by a Tsunami, Partially flooded high density region with diverse building sizes, suburban region with near uniform building sizes and moderate density with localized housing damage due to forest fires, similar suburban region with widespread damage due to forest fires, and finally a rural area with small and sparse buildings. The baseline network has a large number of false positives and false negatives in the examples with significant numbers of buildings. **Segmentation mask colors:** Blue (areas outside buildings), green (undamaged buildings), red (completely destroyed buildings), pink (buildings with major damage) orange (buildings with minor damage). Best viewed in high resolution color.

C. Ablation Studies

In order to study the impact of our design choices, we carry out ablation studies with different loss functions, segmentation heads and with or without the change detection head. The result for each of these experiments are discussed in this section.

1) *Losses*: The impact of using different loss functions can be observed from Table II. Using the localization aware loss provides a substantial improvement over the standard cross-entropy loss and adding in the Dice Loss provides a further, smaller boost.

TABLE II
ABLATION STUDIES

Loss Function	Score
Cross-Entropy Loss	0.69
Localization Aware Loss	0.75
Localization Aware Loss + Dice Loss	0.77
Segmentation Head Architecture	
Simple (Convolution+Upscaling)	0.74
Encoder-Decoder	0.77
Change Detection Head	
Without change detection head	0.76
With change detection head	0.77

2) *Segmentation Head Architecture*: Using the Encoder-Decoder segmentation head provides a boost in performance over the simple segmentation head, as can be seen in Table II.

3) *Change Detection Head*: Finally, we can separate out the impact of using the change detection head which provides a boost of about 1 percentage point. See Table II.

D. Generalization

In order to verify the model's ability to generalize across different geographical regions and disaster types, we carry out a couple of additional experiments where the model is trained on Tier 1 training data and tested on Tier3 data, which belongs to a different set of disasters, and hence is a good measure of generalization ability. The results for these experiments can be seen in Table III. The results indicate that the model is able to generalize on the building segmentation task, but suffers significant degradation on the damage assessment task.

TABLE III
GENERALIZATION ACROSS REGIONS AND DISASTER TYPES

Train Set	Test Set	Score		
		Localization	Damage	Overall
Tier1 Train	Tier1 Valid.	0.79	0.60	0.66
Tier1 Train	Tier3	0.77	0.37	0.50

E. Comparison

Baseline Results for xBD dataset are reported in [1]. As previously noted, this is a two stage approach, where the first stage is a U-Net based building segmentation model and the second stage is two-stream damage classification model that operates on building tiles. Results reported in the baseline Paper are for the Tier 1 test set, which is not available to us

at present, hence for a fair comparison we train the baseline with our dataset split using code provided by the authors and using the same hyper parameters.

Table IV provides a comparison of the overall score achieved by our model against the score achieved by the baseline model. Results obtained by training the baseline building segmentation model and damage classification model are shown in Table V along with results reported in the original paper, as can be seen, we are able to closely reproduce the baseline results. Classwise metrics for the baseline damage classification model are compared with our results in Table VI. RescueNet is able to correctly classify the intermediate damage levels an order of magnitude better than the baseline.

TABLE IV
OVERALL RESULTS (XVIEW2 METRIC)

Model	Localization Score	Damage Score	Overall Score
Ours	0.84	0.74	0.77
Baseline [1]	0.79	0.03	0.26

TABLE V
RESULTS FOR COMPONENT NETWORKS OF BASELINE MODEL

Metric	Baseline (Our Split)	Baseline Reported in [1]
Building segmentation network		
Average IoU	0.65	0.66
Damage Classification network		
Weighted F1	0.6274	0.2654
Macro-Averaged F1	0.3100	0.3204
Harmonic Mean F1	0.0761	0.0342

TABLE VI
CLASS-WISE DAMAGE CLASSIFICATION F1-SCORES

Damage class	Baseline (Our Split)	RescueNet (Ours)
undamaged	0.7211	0.8832
minor	0.0235	0.5628
major	0.0105	0.7711
destroyed	0.4262	0.8079
Harmonic Mean	0.0282	0.7348

VI. CONCLUSION

In this paper, we described a novel unified model for simultaneously segmenting buildings and assessing damage level caused to them by natural disasters using multi-temporal satellite imagery. We achieved this by designing a multi-headed architecture and hierarchical loss function suited to the composite task. We carried out ablation studies to quantify the impact of our loss functions, segmentation head architecture and change detection head, and showed that it was able to generalize across geographic regions and disaster types, while significantly outperforming the existing baseline. For future work on this problem, we plan to focus on better techniques for resolving separate building instances, which would help improve both building localization and damage assessment.

ACKNOWLEDGMENT

The authors would like to thank the organizers of the XView2 Challenge, which first brought this problem to our attention.

REFERENCES

- [1] 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," 2019.
- [2] L. Duffy. Fighting natural disasters with satellite imagery. [Online]. Available: <https://www.pobonline.com/articles/101492-fighting-natural-disasters-with-satellite-imagery>
- [3] R. Foulser-Piggott, R. Spence, K. Saito, D. Brown, and R. Eguchi, "The use of remote sensing for post-earthquake damage assessment: lessons from recent events, and future prospects."
- [4] C. Axel and J. A. van Aardt, "Building damage assessment using airborne lidar," *Journal of Applied Remote Sensing*, vol. 11, no. 4, p. 046024, 2017.
- [5] J. Doshi, D. Garcia, C. Massey, P. Lluca, N. Borensztein, M. Baird, M. Cook, and D. Raj, "Firenet: Real-time segmentation of fire perimeter from aerial video," *Artificial Intelligence for Humanitarian Assistance and Disaster Response Workshop, NeurIPS*, 2019. [Online]. Available: <https://arxiv.org/abs/1910.06407>
- [6] G. Mateo-Garcia, S. Oprea, L. Smith, J. Veitch-Michaelis, A. G. Baydin, and D. Backes, "Flood detection on low cost orbital hardware," *Artificial Intelligence for Humanitarian Assistance and Disaster Response Workshop, NeurIPS*, 2019. [Online]. Available: <https://arxiv.org/abs/1910.03019>
- [7] A. J. Cooner, Y. Shao, and J. B. Campbell, "Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 haiti earthquake," *Remote Sensing*, vol. 8, no. 10, p. 868, 2016.
- [8] F. Nex, D. Duarte, F. G. Tonolo, and N. Kerle, "Structural building damage detection with deep learning: Assessment of a state-of-the-art cnn in operational conditions," *Remote sensing*, vol. 11, no. 23, p. 2765, 2019.
- [9] Y. Li, W. Hu, H. Dong, and X. Zhang, "Building damage detection from post-event aerial imagery using single shot multibox detector," *Applied Sciences*, vol. 9, no. 6, p. 1128, 2019.
- [10] J. Z. Xu, W. Lu, Z. Li, P. Khaitan, and V. Zaytseva, "Building damage detection in satellite imagery using convolutional neural networks," *Artificial Intelligence for Humanitarian Assistance and Disaster Response Workshop, NeurIPS*, 2019. [Online]. Available: <https://arxiv.org/abs/1910.06444>
- [11] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, "Deepglobe 2018: A challenge to parse the earth through satellite images," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.
- [12] A. Van Etten, D. Lindenbaum, and T. M. Bacastow, "Spacenet: A remote sensing dataset and challenge series," *arXiv preprint arXiv:1807.01232*, 2018.
- [13] D. Marcos, D. Tuia, B. Kellenberger, L. Zhang, M. Bai, R. Liao, and R. Urtasun, "Learning deep structured active contours end-to-end," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8877–8885.
- [14] I. W. II/4. Isprs 2d semantic labeling contest. [Online]. Available: <http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html>
- [15] 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*. Springer, 2015, pp. 234–241.
- [16] V. Iglovikov, S. Seferbekov, A. Buslaev, and A. Shvets, "Ternausnetv2: Fully convolutional network for instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 233–237.
- [17] R. Hamaguchi and S. Hikosaka, "Building detection from satellite imagery using ensemble of size-specific detectors," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2018, pp. 223–2234.
- [18] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE Visual Communications and Image Processing (VCIP)*. IEEE, 2017, pp. 1–4.
- [19] S. Golovanov, R. Kurbanov, A. Artamonov, A. Davydov, and S. Nikolenko, "Building detection from satellite imagery using a composite loss function," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 229–232.
- [20] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: a systematic survey," *IEEE transactions on image processing*, vol. 14, no. 3, pp. 294–307, 2005.
- [21] J. Zhao, M. Gong, J. Liu, and L. Jiao, "Deep learning to classify difference image for image change detection," in *2014 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2014, pp. 411–417.
- [22] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 4063–4067.
- [23] R. Caye Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Guided anisotropic diffusion and iterative learning for weakly supervised change detection," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.
- [24] M. Papadomanolaki, S. Verma, M. Vakalopoulou, S. Gupta, and K. Karantzas, "Detecting urban changes with recurrent neural networks from multitemporal sentinel-2 data," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 214–217.
- [25] 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*, 2015, pp. 91–99.
- [26] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.
- [27] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [29] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 472–480.
- [30] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.
- [31] R. Deng, C. Shen, S. Liu, H. Wang, and X. Liu, "Learning to predict crisp boundaries," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 562–578.