

# RescueNet: A High Resolution UAV Semantic Segmentation Benchmark Dataset for Natural Disaster Damage Assessment

Tashnim Chowdhury, *Member, IEEE*, Robin Murphy, *Fellow, IEEE*, and Maryam Rahnemoonfar, *Member, IEEE*

**Abstract**—Due to climate change, we can observe a recent surge of natural disasters all around the world. These disasters are causing disastrous impact on both nature and human lives. Economic losses are getting greater due to the hurricanes. Quick and prompt response of the rescue teams are crucial in saving human lives and reducing economic cost. Deep learning based computer vision techniques can help in scene understanding, and help rescue teams with precise damage assessment. Semantic segmentation, an active research area in computer vision, can put labels to each pixel of an image, and therefore can be a valuable arsenal in the effort of reducing the impacts of hurricanes. Unfortunately, available datasets for natural disaster damage assessment lack detailed annotation of the affected areas, and therefore do not support the deep learning models in total damage assessment. To this end, we introduce the RescueNet, a high resolution post disaster dataset, for semantic segmentation to assess damages after natural disasters. The RescueNet consists of post disaster images collected after Hurricane Michael. The data is collected using Unmanned Aerial Vehicles (UAVs) from several areas impacted by the hurricane. The uniqueness of the RescueNet comes from the fact that this dataset provides high resolution post-disaster images and comprehensive annotation of each image. While most of the existing dataset offer annotation of only part of the scene, like building, road, or river, RescueNet provides pixel level annotation of all the classes including building, road, pool, tree, debris, and so on. We further analyze the usefulness of the dataset by implementing state-of-the-art segmentation models on the RescueNet. The experiments demonstrate that our dataset can be valuable in further improvement of the existing methodologies for natural disaster damage assessment.

**Index Terms**—Semantic segmentation, UAV, natural disaster damage assessment.

## I. INTRODUCTION

In recent year, due to climate change and other reasons, numerous natural disasters have affected different parts of the world. The impacts of these disasters are getting stronger and more lasting. Reducing economic losses and saving valuable human lives depends heavily on quick response from the rescue teams. Different compute vision technique can assist in accurate damage assessment exploring the visual components of an imagery. Among different fields of computer vision, semantic segmentation is one of the essential parts of computer vision. Semantic segmentation aims to classify each pixel of an image. Semantic segmentation can classify each object of an image

Tashnim Chowdhury and Maryam Rahnemoonfar are with the Department of Information Systems, University of Maryland, Baltimore County, Baltimore, MD, 21250 USA.

Robin Murphy is with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843 USA

Corresponding Author: Maryam Rahnemoonfar, e-mail: maryam@umbc.edu.

with distinct boundaries. Recently, with the application of deep learning methodologies, significant advances have been made in scene understanding of urban scenes. Several pioneering datasets including Cityscapes [1], PASCAL VOC2012 [2], PASCAL Context [3], and COCO Stuff dataset [4] are available and these large-scale datasets are helping in improvement of segmentation accuracy of urban scenes. Despite these advances in semantic segmentation, it still remains a challenging task in natural disaster damage assessment due to the lack of benchmark datasets.

For natural disaster damage assessment there have been several datasets proposed in recent years. Most of the datasets are either satellite dataset [5], [6] or datasets collected from social media images [7]–[9]. Research community of this area are currently facing two types of difficulties: 1) lack of comprehensive pixel level annotation of post-disaster scenes, and 2) classification of damage labels. The first issue arises due to lack of comprehensive annotation of whole scene. There is a lack of datasets that provide pixel level annotation. Besides these few datasets only provide annotation of few classes. For example, [10] provides semantic labels, and [6], [11] provide bounding box annotation of different buildings. Damage information of few classes does not provide complete scene understanding. Other elements of the images including road, vehicle, damaged trees, debris can explain the scene extensively, and help in more accurate damage assessment. Although FloodNet [12] includes pixel level annotation of different components of post disaster images including above mentioned classes like vehicle, road, and trees, it lacks annotation of damaged buildings and debris. To resolve this issue, in this paper, we introduce a low altitude and high resolution natural disaster dataset named RescueNet. RescueNet provides pixel level annotation of 11 distinct categories including debris, water, building, vehicle, road, tree, pool, and sand. The second issue is mainly related to classification of different damage levels of a category. For example, after a hurricane or wildfire, a building or road can be damaged. But how much damage is done on that particular structure is not mentioned in most of the datasets. Moreover, a building can be damaged in different ways. It can be slightly damaged, or it can be totally destroyed after a disaster. Although [6] classify the buildings based on four different damage levels, this is low resolution satellite dataset, and therefore the image quality is poor. To this end, our proposed RescueNet provides building segmentation labels with four different damage criteria where images are very high in resolution.

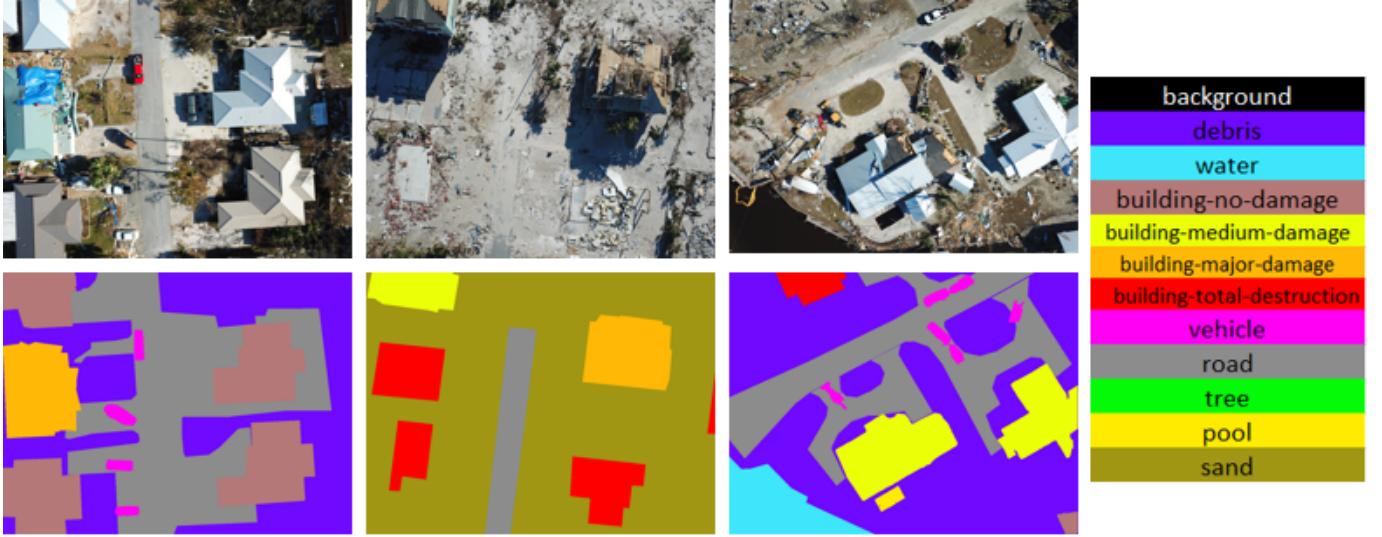


Fig. 1. Illustration of complex scenes of RescueNet dataset. First row shows original image, and the second row shows the corresponding annotations.

RescueNet is a high resolution aerial dataset collected using UAVs (Unmanned Aerial Vehicle). It provides high resolution images of the affected areas after hurricane Michael. Besides creating different classes of different objects, more classes have been created to distinguish damage levels of different objects. RescueNet consists of 4494 images of both affected and non-affected areas after the hurricane. It provides pixel level annotation of in total 11 classes including debris, water, building-superficial-damage, building-medium-damage, building-major-damage, building-total-destruction, vehicle, road, pool, tree, and sand. Few examples of RescueNet and their corresponding annotations are shown in Figure 1. To the best of our knowledge, this is the only dataset that provides pixel level annotation of debris accompanied with several other components of an affected area. Finally, we evaluate the performance of state-of-the-art semantic segmentation methods on RescueNet. The experiments indicate the limitation of the current state-of-the-art methods and highlights the challenges of existing methods in generalizing the actual after disaster scenarios.

In summary the contributions of this paper are as follows:

- We present RescueNet, a high resolution aerial imagery, captured using UAVs.
- We provide high quality annotation of 11 classes which includes images from both affected and non-affected areas after hurricane Michael.
- We also provide annotation of building damages based on the severity. The building damages are classified into four damage classes such as superficial damage, medium damage, major damage, and total destruction.
- We experiment with several existing semantic segmentation methods on RescueNet and show the usefulness of the dataset in future research of natural disaster damage assessment.

## II. RELATED WORKS

### A. Existing Natural Disaster Damage Assessment Dataset

Existent natural disaster datasets can be sorted into two classes. One is ground-level images [7]–[9], and other one is satellite and aerial imagery [5], [6]. The ground level images are collected from different sources including Virtual Disaster Viewer [8], Open Images Dataset [8], Google image search engine [8], [9], and social networks [7], [13]. Authors in [14] presents a large ground level post disaster dataset for different incident detection like drought, wildfire, and snowstorm. Images has been introduced to the AIDR system [13] by Nguyen *et al.* in [7] by collecting images from the social media. Although these imageries are abundant, they lack geo location tags [11].

On the other hand, satellite and aerial imagery is collected from different satellites and remote sensing equipment. A post tsunami aerial image dataset named ABCD (AIST Building Change Detection) has been proposed in [15]. This dataset helps in detection of flood affected buildings. Authors in [16] presents a dataset that combines SpaceNet [17] and DeepGlobe [18]. In order to assess the damages caused by hurricanes, Chen *et al.* proposed a dataset in [5] by collecting images from two different sources including crowdsourced annotated data of DigitalGlobe satellite imagery and data collected by FEMA. This dataset provides bounding box annotations of buildings, labeled as either non-damaged or damaged, for object detection. For land and building detection fMoW (Functional Map of the World) is introduced in [19] which consists of around one million images containing bounding box annotations of 63 categories. Rudner *et al.* in [10] present a dataset for semantic segmentation of flooded buildings where images are collected from two satellites named Sentinel-1 and Sentinel-2. A large dataset named xBD which consists of both pre- and post-disaster satellite images are proposed by Gupta *et al.* in [6] for building damage assessment with 4 different damage categories, No Damage, Minor Damage, Major Damage, and Destroyed. xBD provides satellite imagery from a variety of disaster events

with building polygons, classification labels for damage types, ordinal labels of damage level, and corresponding satellite metadata. ISBDA (Instance Segmentation in Building Damage Assessment) is presented by Zhu *et al.* in [11]. This dataset includes user-generated aerial videos collected from social media platform and annotated with 3 instance-level building damage mask including Slight, Severe, and Debris. One of the main drawbacks of satellite imageries is that the images are low in resolution. If someone is interested in detailed damage of different structures like roads, and buildings, images with higher resolution are required which is lacking in satellite imageries. Compared to satellite imagery RescueNet is an UAV based imagery which provides high resolution images which help in detailed damage assessment.

Besides satellite imagery, very few UAV based disaster datasets are also available. Kyrou *et al.* propose an image classification dataset named AIDER (Aerial Image Database for Emergency Response) in [20]. AIDER consists of images from 4 different disaster events including Fire/Smoke, Flood, Collapsed Building/Rubble, and Traffic Accidents. Rahmehoonfar *et al.* present a high resolution post hurricane dataset named FloodNet in [12]. This dataset provides pixel level annotation of 9 classes for natural disaster damage assessment. The UAV images provided by FloodNet consists of flooded and non-flooded areas collected after hurricane Harvey. Although RescueNet is similar to FloodNet, they are different in many aspects. While FloodNet addresses the flooded buildings and roads by classifying them into 2 classes (flooded, non-flooded), RescueNet classify building damages into 4 classes (no-damage, medium-damage, major-damage, total-destruction). The most prominent difference is that RescueNet provides pixel level annotation of “debris”, which is not present in FloodNet dataset since there is no damaged building. Both pixel-level annotation and inclusion of all objects in RescueNet will certainly help to estimate post disaster damages.

Moreover, compared to existing ground-level and aerial/satellite imagery, RescueNet provides semantic segmentation of all the objects (building, road, tree, pool, debris, sand etc.) present in the images compared to other datasets which provide only a part (building, road) of the whole image. The images of RescueNet are collected using UAVs which provides high resolution compared to other satellite datasets [6]. A comparative study of different existing natural disaster datasets with RescueNet is presented in Table I.

### B. Natural Disaster Damage Assessment Methodologies

In recent times several research works have been proposed for the assessment of damages caused different natural disasters. River segmentation is performed in [21] to monitor flood water using three existing segmentation methods [22]–[24]. Doshi *et al.* perform semantic segmentation on satellite images in [16] in order to detect impact of natural disasters on man-made structures. Rahmehoonfar *et al.* in [25] present a recurrent neural network for flood detection using UAV images. Multi3Net [10] is proposed by Rudner *et al.* for segmentation of flooded buildings using satellite imagery. Chowdhury *et al.*

present a self-attention based semantic segmentation method in [26], and another in [27] for post disaster damage assessment after hurricane Michael. Performance of existent segmentation methods are also evaluated on aerial images by Gupta *et al.* in [28] for building and road segmentation. A multilevel instance segmentation method named MSNet is proposed by authors in [11] which assesses building damages.

Although several semantic segmentation methods have been implemented and proposed on different natural disaster datasets, there is no clear discussion on which type of segmentation networks perform accurately and efficiently. In this paper, we experiment with four segmentation networks from different genre types, and discuss their performance in terms of accuracy and efficiency.

### C. Semantic Segmentation

Semantic segmentation is one of the prime research area in computer vision. With advancement of deep learning several recent methods have improved the performance of semantic segmentation on images and videos. A pioneering work in this field, fully convolutional networks (FCN) [22], inspires the researchers to explore the deep learning methods in computer vision. Existing segmentation methods can be classified into two classes: non-attention [29]–[31] and attention based methods [26], [27], [32]–[35]. Non-attention based methods can be further classified into encoder-decoder [31] and pyramid pooling based methods [29], [30].

Encoder-decoder based methods like U-Net [31] generate feature maps using local contexts collected from middle and lower level features through encoder-decoder architecture. This type of architectures can pick small details and produce sharp object boundaries. Thus, the prediction from these methods are high in resolution. On the other hand, pyramid based modules [29], [30] use global average pooling, pyramid pooling [29], and atrous convolution [30] to generate feature maps using global contexts. Pyramid pooling based methods can generate high resolution predictions at different scales.

Attention modules can model long range dependencies and are being used in several applications such as visual question answering, video classification, and machine translation. Although introduced first in machine translation problem [36], self-attention based methods soon has gained popularity in computer vision. From computer vision perspective, self-attention calculates the context at one position as weighted sum of all positions in a sentence or an image. Recently several self-attention based techniques [32], [33], [37] have been introduced and these techniques have shown promising performances in different tasks including semantic segmentation. Authors in [33] propose a dual attention network which uses both channel and position attention to capture interdependencies of features among spatial and channels dimensions. Huang *et al.* proposes a criss-cross attention module to gather contextual information in spatial domain. Object based context is introduced in [37].

These state-of-art semantic segmentation networks have been mainly applied on ground based imagery [1], [3]. Although few research works [21], [28] have applied few popular network models on both satellite and aerial imageries, these works do

TABLE I  
OVERVIEW OF EXISTING NATURAL DISASTER DATASETS.

Dataset	Size	Resolution	Image Type	Task	# of Annotated Classes
ABCD [15]	22171	Varies	Satellite	Classification	2
Chen <i>et al.</i> [5]	-	Varies	Satellite	Object Detection	2
fMoW [19]	~ 1 million	Varies	Satellite	Classification	63
Rudner <i>et al.</i> [10]	-	Varies	Satellite	Semantic Segmentation	2
xBD [6]	22068	1024x1024	Satellite	Instance Segmentation, Classification	4
ISBDA [11]	1030	-	Aerial (Social Media)	Object Detection	3
AIDER [20]	2545	-	UAV	Classification	4
FloodNet [12]	2343	3000x4000	UAV	Semantic Segmentation	9
<b>RescueNet</b>	<b>4494</b>	<b>3000x4000</b>	<b>UAV</b>	<b>Semantic Segmentation</b>	<b>11</b>

not include a complete semantic segmentation of these images. In contrast to other research works, we apply four state-of-art semantic segmentation network models on our proposed RescueNet dataset for a complete scene segmentation. We adopt one encoder-decoder based network named ENet [38], one pyramid pooling module based network PSPNet [29], and the last network model DeepLabv3+ [30] employs both encoder-decoder and pyramid pooling module. Besides these non-attention based methods we also adopt an attention based network called Attention U-Net [34].

### III. DATASET

#### A. Data Collection

The dataset is taken from the Center for Robot-Assisted Search and Rescue open data repository ([hrail.crasar.org](http://hrail.crasar.org)) for small UAV (sUAV) imagery for disasters, specifically the Hurricane Michael event. The two most important features of this dataset are: fidelity and uniqueness. First of all, this dataset is collected by the emergency responders during the response phase using sUAV. This reflects the current practice of data collection and can be expected to be collected during a disaster. Secondly, it is unique since it is the only known database of sUAV imagery for disaster. Note that there is other existing database of imagery collected using unmanned and manned aerial assets during different disasters, such as National Guard Predators or Civil Air Patrol. But compared to our dataset, those are collected using larger and fixed-wing assets that operate above 400 feet AGL (above ground level) limitation of sUAV.

Hurricane Michael made landfall near Mexico Beach, Florida, on October 10, 2018, as a category 5 hurricane, one of the powerful and destructive tropical cyclones to strike United States since Andrew in 1992. The dataset was collected on behalf of Florida State Emergency Response Team at Mexico Beach and other directly impacted areas [39] after operating 80 flights conducted between October 11-14, 2018. Unlike typical manned assets which normally fly at 500 feet AGL, the flights were flown at 200 feet AGL. While images were taken using DJI Mavic Pro quadcopters, two sets of videos were taken with Parrot Disco fixed-wing sUAV and one set at night with a DJI Inspire and thermal camera. But these videos are not included in the RescueNet.

#### B. Data Annotation

V7 Darwin platform [40] is used to annotate the dataset for semantic segmentation. With a goal of complete pixel level annotation of an image, we annotate all the objects present in the dataset which consists of debris, water, building, road, tree, pool, and sand. The incentive of including all objects for pixel level annotation is that identifying all damaged, and non-damaged objects in an image helps in providing a better understanding of the actual damage done by a natural disaster. The building class includes both residential, and non-residential structures. Following the FEMA guideline [41], we further classify the building damages into four folds: Building-No-Damage, Building-Medium-Damage, Building-Major-Damage, and Building-Total-Destruction. A summary of annotated polygons of these four classes is shown in Table II.

TABLE II  
NUMBER OF POLYGONS OF DIFFERENT BUILDINGS BASED ON THEIR DAMAGE LEVELS.

Damage Level	Number of Polygons
No Damage	4011
Medium Damage	3119
Major Damage	1693
Total Destruction	2080

To define different damage levels of buildings, we follow the guideline provided by the FEMA [41]. Although the guideline includes the scenario where the buildings conditions can be visible from all sides, aerial images provide only top views. Therefore, we have to adapt to the definitions based on the top views. According to the first responders, the building damage classes are defined as follows. If no damage is done to a building, then the building is classified as Building-No-Damage. On the other hand, if some parts of a building are affected, and requires minimal repairs (for example, the roof can be covered with a blue tarp) to make it habitable, then the building is classified as Building-Medium-Damage. However, if the damage is severe enough that it has sustained significant structural damage, and in need of extensive repairs, then the building is denoted as Building-Major-Damage. And when two or more major structural components of a building collapse

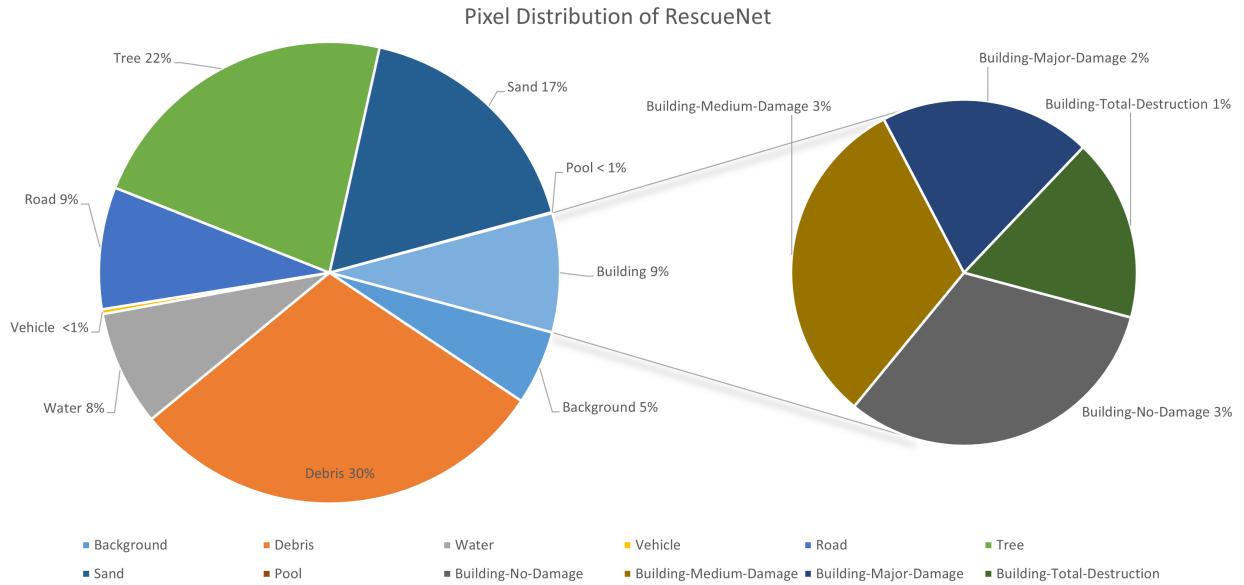


Fig. 2. Pixel distribution of different classes in RescueNet.

(for example, collapse of basement walls, foundation, load-bearing walls, or roof), then the building is categorized as Building-Total-Destruction.

Few examples from RescueNet and their colored annotated masks are shown in Figure 1. In the first image, some damaged buildings, debris, and water are visible. The second image shows some totally destroyed buildings, debris, and cars. It is apparent from this image is that distinguishing between debris and totally destroyed building is difficult due their similar texture. Most part of the road is also almost invisible due to sand and debris. The third image shows another damaged area with some building and lots of debris.

First responders are not only interested in damaged roads and buildings, but they also want an estimate of debris in an area. This will help them to allocate relevant machineries to remove debris from an area. Since it is not possible to annotate each single debris, for any image we annotate the whole ground as debris if there is a significant amount of debris.

### C. Quality Control

To ensure the quality of the annotation, each image passes through two steps verification process. A guideline was provided to the annotators which includes the definition of different classes and features to distinguish among different building damages as mentioned in section III-B. Annotation of each image takes about an hour since we did pixel level annotation all the objects of the image. For the quality control each image initially passed to an annotator. After the annotation is done, the image is sent to a reviewer. The reviewer verifies the quality of the annotation to make sure that the annotation is good in quality and is accordance to the definition provided in the guideline. If the annotation is accurate the annotated image is accepted, otherwise it is sent back to the annotator with detailed comments for further correction. This cycle of

review and correction continues until the reviewers are satisfied with quality of the annotations. By this way we confirm the good quality of the annotation of the whole dataset.

### D. Dataset Statistics

RescueNet is collected from the hurricane Michael affected area using UAVs. The imagery includes several classes including debris, water, building-no-damage, building-minor-damage, building-major-damage, building-total-destruction, vehicle, road, tree, pool, and sand. One of the prime features of RescueNet is the extended semantic labels of buildings based on their damage levels. As shown in Table II the dataset has 4011 polygons of building with no damage labels, 3119 polygons of building with minor damage, 1693 polygons of building with major damage, and 2080 polygons of building with total destruction. Pixel distribution of different classes in the RescueNet is shown in Figure 2.

### E. Dataset Splits

We split our dataset into three subsets, training, validation, and testing. But the dataset contains both disasters affected areas and non-affected areas. In some images, areas are covered with debris and lots of buildings of different damage labels (superficial, medium, major, total destruction) are present. Therefore, to ensure uniform distribution of images of both affected and non-affected areas throughout the three subsets we classify the dataset into three folds: Superficial damage, Medium damage, and Major damage. We consider three classification labels for the images based on how much damage is visible in the area covered by an image. If the area presented by an image does not show any damaged structures (man-made and natural), then this image is classified as “Superficial damage”. If few structures are damaged by the natural disaster, then the image is classified as “Medium damage”. Finally,

if at least one totally destroyed building or around 50% area is covered with debris, then the image is labeled as “Major Damage”. From each category, 80% of the images are distributed to training set, 10% to the validation, and 10% to the test set.

#### F. Unique Dataset Features

RescueNet provides several unique features compared to other existing natural disaster datasets. Although there are several natural disaster datasets for image classifications, there are not much available dataset available for semantic segmentation. RescueNet provides a dataset that contains comprehensive semantic labels of all the objects present in every images. The only another dataset that provides this kind of complete semantic labels of each objects for natural disaster damage assessment is FloodNet [12]. But FloodNet does not have any damaged building and there is no “debris” class which is unique to RescueNet.

Besides RescueNet is unique from the perspective of semantic labels of class “debris”. Most of the dataset include damage of different objects like building and roads. But they do not include “debris”, which is a crucial part after any natural disaster. Specially for residential areas determination of spreading of debris and amount of debris can help the rescue team to send proper machinery to remove the debris.

Another unique feature of RescueNet is the different classification of building damage levels. Although few datasets like xBD [6] provide different levels of building damages, they provide instance level semantic segmentation annotations from satellite imageries which are low in resolution. RescueNet not only provides pixel level annotation to the buildings based on different damage levels but also the images are higher in resolution. The higher resolution helps in detecting detailed damages inflicted on the buildings.

#### IV. POTENTIAL USES OF RESCUENET

There are several usages of RescueNet which are both important from practical point of view and from the perspective of academic research.

**Damage classification of buildings.** RescueNet provides building damages with explicit classification of damage level. This information is crucial to first responders and rescue planners to take accurate decision regarding allocate their rescue efforts.

**Assessment of debris.** An accurate assessment of amount of debris in an affected area is also imperative during and after a natural disaster. Depending on the amount of debris more physical and heavy machinery is required to remove the debris. Otherwise, the rescue effort will be hampered and might cost precious human lives.

**Segmentation of small objects.** RescueNet comes with a very practical scenario where some objects are comparatively smaller than others. This dataset includes images of pools and vehicles which are significantly smaller than buildings and roads. Segmentation of smaller objects is an active research area, and this dataset will help the researchers to design models that will segment smaller objects more efficiently and implement those models in natural disaster damage assessment.

#### V. BENCHMARK FOR SEMANTIC SEGMENTATION

**Problem Details.** The objective of semantic segmentation, one of the fundamental tasks in computer vision, is to put label to each pixel of an image. RescueNet consists of post disaster images and provide pixel level annotation of 11 object classes. For any image in the dataset, the RescueNet has annotation of all objects present in the scene along with detailed damage levels of some of the classes such as buildings, and roads.

**Models and Training Details.** We implement four state-of-the-art semantic segmentation methods namely ENet [38], PSPNet [29], DeepLabv3+ [30], and Attention U-Net [34] and evaluated their performance on RescueNet. We implement the methods using PyTorch and use NVIDIA GeForce RTX 2080 Ti GPU and Intel Core i9 CPU as hardware. We use resnet101 as backbone for PSPNet and DeepLabv3+. We implement “poly” learning rate where base learning rate is 0.001. All the models use the following hyperparameters settings. Momentum is set to 0.9, weight decay to 0.00001, power to 0.9, and weight of the auxiliary rate to 0.4. For data augmentation we implement scaling, flipping, random shuffling, and random rotation. Data augmentation helps in avoiding overfitting. We resize the images to  $713 \times 713$  during training. We use mean Intersection over Union (mIoU) as evaluation metric for semantic segmentation.

**Benchmark.** Table III shown segmentation results of four state-of-art semantic segmentation methods on RescueNet. Among these four methods three (ENet, PSPNet, DeepLabv3+) are non-attention methods and Attention U-Net is an attention method. From the table it is evident that Attention U-Net performed best in terms of mIoU with a score of 93.98% which is 35.57% greater than its closest competitor DeepLabv3+.

#### VI. DISCUSSION

RescueNet poses several challenges because of its small classes present in the dataset. Small classes like “vehicle” and “pool” makes it difficult to get a good segmentation compared to larger objects like buildings and roads. Besides that, since UAV images include only the top view of a scene, it is difficult to assess the actual damage, since horizontal view also brings information regarding all sides of a building. Another prime challenge is “debris”. Fallen debris on roads and water makes it difficult to identify the roads and water, since “debris” itself is another class in the dataset. Similar texture of “debris” and “sand” might also be an issue in segmenting these two classes since the models can be confused in differentiating between these two. Several other challenges include differentiating between trees and debris, and vehicle and debris. When both damaged trees and vehicles are mixed with debris, we consider both as part of the debris. Hence, segmentation model has to understand the context of the scene to accurately segment small objects and objects that might be covered with debris.

From the results obtained four segmentation methods, ENet [38], PSPNet [29], DeepLabv3+ [30], and Attention U-Net [34], it can be seen that attention based methods performed significantly better than non-attention based methods on RescueNet in all categories. Among non-attention based method, DeepLabv3+ performs best compared to ENet and PSPNet. It is surprising

TABLE III  
PER-CLASS RESULTS ON RESCUENET TESTING SET.

Method	Debris	Water	Building No Damage	Building Minor Damage	Building Major Damage	Building Total Destruction	Vehicle	Road	Tree	Pool	Sand	<b>Mean IoU %</b>
ENet [38]	51.03	63.38	46.79	37.43	32.57	39.24	36.21	61.90	70.05	43.29	65.11	49.72
PSPNet [29]	66.04	60.46	50.78	50.60	62.70	58.15	22.75	69.04	73.13	9.92	80.67	54.93
DeepLabv3+ [30]	60.30	75.60	59.10	43.50	43.70	52.70	46.90	69.70	77.00	42.00	72.00	58.41
Attention U-Net [34]	95.32	96.00	96.15	95.93	95.46	93.77	81.89	95.12	97.75	89.11	97.30	93.98

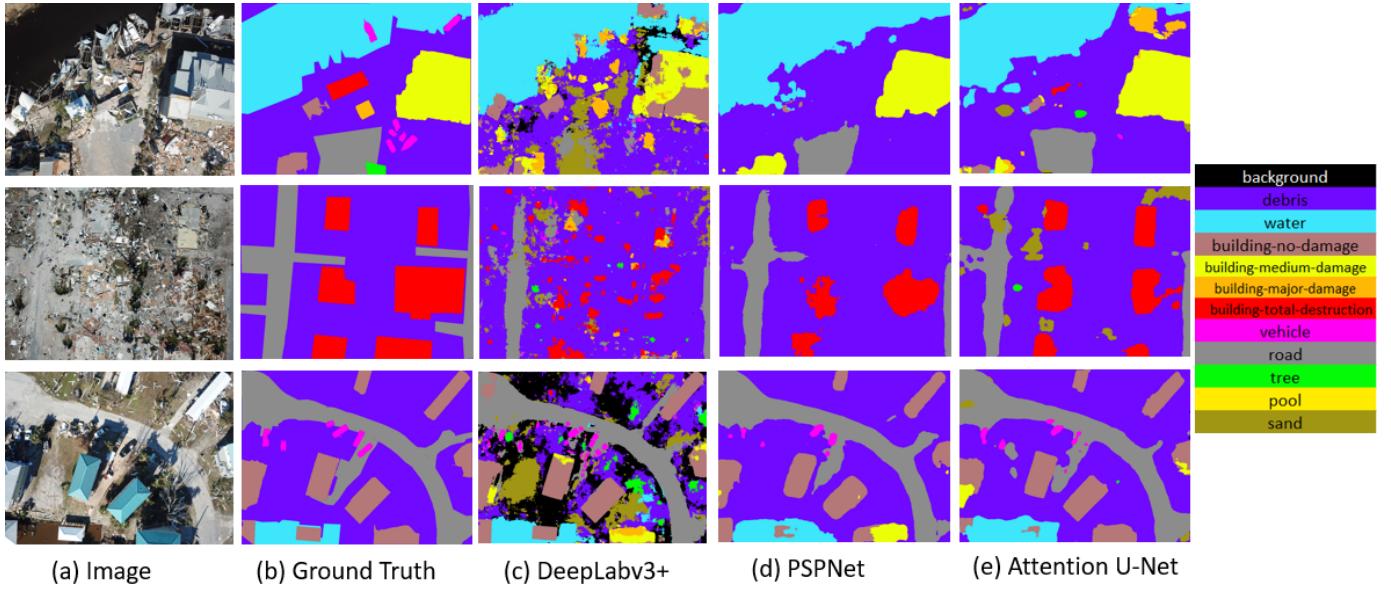


Fig. 3. Visual comparison of PSPNet, DeepLabv3+, and Attention U-Net on RescueNet test set. Although DeepLabv3+ performed better than PSPNet in mean IoU, qualitatively PSPNet generated better predicted images than DeepLabv3+.

to see very low performance of non-attention based methods in all classes specially in small object classes. Specifically, PSPNet achieved only 9.92% in IoU in “pool” class. This is purely a pyramid pooling based method. On the other hand, ENet and DeepLabv3+ are encoder-decoder type segmentation methods. Therefore, from the experiments it is evident that among non-attention based methods, encoder-decoder based methods performed better compared to pyramid pooling based PSPNet in small classes like “vehicle” and “pool”. A further investigation is required to find the probable causes. This can open a new research area to improve segmentation performance of encoder-decoder type architecture on RescueNet. Although Attention U-Net has outperformed all other three methods, this is a very heavy neural network compared to ENet. While ENet has 0.37M model parameters, Attention U-Net has more than 34M parameters. In practical scenario, after a natural disaster first responders would use UAVs to investigate the damage inflicted upon an area. For hardware like cell phones and UAVs, light-weight networks like ENet is the best choice. From the experiments we can observe that ENet does not perform good, actually worst, in semantic segmentation task of the RescueNet. Therefore, there is a great scope of research to improve the performance of the light-weight semantic segmentation network for natural disaster damage assessment.

## VII. CONCLUSION

In this paper, we introduce a high resolution UAV dataset RescueNet which contains comprehensive pixel level annotation of 11 classes for natural disaster damage assessment. We discuss the dataset collection and annotation process along with challenges it poses. We also present dataset statistics for a detailed overview of the dataset and the classes it contains. Four state-of-art semantic segmentation methods have been evaluated on RescueNet and the results are discussed. We hope that this dataset will open new avenues of research on damage management after a natural disaster and will help researchers to proposed improved methods.

## REFERENCES

- [1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.
- [2] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes (voc) challenge,” *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.
- [3] R. Mottaghi, X. Chen, X. Liu, N.-G. Cho, S.-W. Lee, S. Fidler, R. Urtasun, and A. Yuille, “The role of context for object detection and semantic segmentation in the wild,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 891–898.
- [4] H. Caesar, J. Uijlings, and V. Ferrari, “Coco-stuff: Thing and stuff classes in context,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1209–1218.

- [5] S. A. Chen, A. Escay, C. Haberland, T. Schneider, V. Staneva, and Y. Choe, “Benchmark dataset for automatic damaged building detection from post-hurricane remotely sensed imagery,” *arXiv preprint arXiv:1812.05581*, 2018.
- [6] R. Gupta, B. Goodman, N. Patel, R. Hosfelt, S. Sajeev, E. Heim, J. Doshi, K. Lucas, H. Choset, and M. Gaston, “Creating xbd: A dataset for assessing building damage from satellite imagery,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 10–17.
- [7] D. T. Nguyen, F. Ofli, M. Imran, and P. Mitra, “Damage assessment from social media imagery data during disasters,” in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 2017, pp. 569–576.
- [8] K. R. Nia and G. Mori, “Building damage assessment using deep learning and ground-level image data,” in *2017 14th conference on computer and robot vision (CRV)*. IEEE, 2017, pp. 95–102.
- [9] E. Weber, N. Marzo, D. P. Papadopoulos, A. Biswas, A. Lapedriza, F. Ofli, M. Imran, and A. Torralba, “Detecting natural disasters, damage, and incidents in the wild,” in *European Conference on Computer Vision*. Springer, 2020, pp. 331–350.
- [10] T. G. Rudner, M. Rufwurm, J. Fil, R. Pelich, B. Bischke, V. Kopačková, and P. Biliński, “Multi3net: segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 702–709.
- [11] X. Zhu, J. Liang, and A. Hauptmann, “Msnet: A multilevel instance segmentation network for natural disaster damage assessment in aerial videos,” *arXiv preprint arXiv:2006.16479*, 2020.
- [12] M. Rahmehoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. R. Murphy, “Floodnet: A high resolution aerial imagery dataset for post flood scene understanding,” *IEEE Access*, vol. 9, pp. 89 644–89 654, 2021.
- [13] D. T. Nguyen, F. Alam, F. Ofli, and M. Imran, “Automatic image filtering on social networks using deep learning and perceptual hashing during crises,” *arXiv preprint arXiv:1704.02602*, 2017.
- [14] E. Weber, N. Marzo, D. P. Papadopoulos, A. Biswas, A. Lapedriza, F. Ofli, M. Imran, and A. Torralba, “Detecting natural disasters, damage, and incidents in the wild,” in *The European Conference on Computer Vision (ECCV)*, August 2020.
- [15] A. Fujita, K. Sakurada, T. Imaizumi, R. Ito, S. Hikosaka, and R. Nakamura, “Damage detection from aerial images via convolutional neural networks,” in *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*. IEEE, 2017, pp. 5–8.
- [16] J. Doshi, S. Basu, and G. Pang, “From satellite imagery to disaster insights,” *arXiv preprint arXiv:1812.07033*, 2018.
- [17] D. CosmiQWorks, “Nvidia: Spacenet on amazon web services (aws) datasets: The spacenet catalog.”
- [18] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raska, “Deepglobe 2018: A challenge to parse the earth through satellite images,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2018, pp. 172–17209.
- [19] G. Christie, N. Fendley, J. Wilson, and R. Mukherjee, “Functional map of the world,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6172–6180.
- [20] C. Kyrou and T. Theocharides, “Deep-learning-based aerial image classification for emergency response applications using unmanned aerial vehicles,” in *CVPR Workshops*, 2019, pp. 517–525.
- [21] L. Lopez-Fuentes, C. Rossi, and H. Skinnemoen, “River segmentation for flood monitoring,” in *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 2017, pp. 3746–3749.
- [22] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [23] S. Jégou, M. Drozdzal, D. Vázquez, A. Romero, and Y. Bengio, “The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 11–19.
- [24] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [25] M. Rahmehoonfar, R. Murphy, M. V. Miquel, D. Dobbs, and A. Adams, “Flooded area detection from uav images based on densely connected recurrent neural networks,” in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 1788–1791.
- [26] T. Chowdhury and M. Rahmehoonfar, “Attention based semantic segmentation on uav dataset for natural disaster damage assessment,” in *2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2021, pp. 2325–2328.
- [27] ———, “Self attention based semantic segmentation on a natural disaster dataset,” in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 2798–2802.
- [28] A. Gupta, S. Watson, and H. Yin, “Deep learning-based aerial image segmentation with open data for disaster impact assessment,” *arXiv preprint arXiv:2006.05575*, 2020.
- [29] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.
- [30] 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.
- [31] 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.
- [32] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, “Ccnnet: Criss-cross attention for semantic segmentation,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 603–612.
- [33] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, “Dual attention network for scene segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3146–3154.
- [34] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, “Attention u-net: Learning where to look for the pancreas,” *arXiv preprint arXiv:1804.03998*, 2018.
- [35] T. Chowdhury and M. Rahmehoonfar, “Attention for damage assessment,” in *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*, 2021.
- [36] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” *arXiv preprint arXiv:1706.03762*, 2017.
- [37] Y. Yuan and J. Wang, “Ocnet: Object context network for scene parsing,” *arXiv preprint arXiv:1809.00916*, 2018.
- [38] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, “Enet: A deep neural network architecture for real-time semantic segmentation,” *arXiv preprint arXiv:1606.02147*, 2016.
- [39] O. Fernandes, R. Murphy, D. Merrick, J. Adams, L. Hart, and J. Broder, “Quantitative data analysis: Small unmanned aerial systems at hurricane michael,” in *2019 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2019, Conference Proceedings, pp. 116–117.
- [40] “V7 darwin,” <https://www.v7labs.com/darwin>, accessed: 2020-08-25.
- [41] “Fema guideline,” [https://www.fema.gov/sites/default/files/2020-07/fema\\_preliminary-disaster-assessment\\_guide.pdf](https://www.fema.gov/sites/default/files/2020-07/fema_preliminary-disaster-assessment_guide.pdf), accessed: 2021-08-25.



**Tashnim Chowdhury** received his B.S. degree in Electrical and Electronic Engineering from the Chittagong University of Engineering and Technology, Chittagong, Bangladesh in 2013. He also did his M.S. in Electrical Engineering from The University of Toledo, Ohio, USA in 2016. He is currently pursuing his Ph.D. in Information Systems from the University of Maryland Baltimore County, USA. His research interest includes Deep Learning, Machine Learning, Semantic Segmentation, Few Shot Learning, Meta Learning, and Bayesian Learning.



**Robin Roberson Murphy** received a B.M.E. in mechanical engineering, a M.S. and Ph.D. in computer science in 1980, 1989, and 1992, respectively, from Georgia Tech, where she was a Rockwell International Doctoral Fellow. She is the Raytheon Professor of Computer Science and Engineering at Texas A&M and directs the Center for Robot-Assisted Search and Rescue and an IEEE Fellow. Her research interests are artificial intelligence, human-robot interaction, and heterogeneous teams of robots and she has over 100 publications including the best selling textbook, Introduction to AI Robotics (MIT Press 2000). She is a founder of the fields of rescue robots and human-robot interaction.



**Maryam Rahnemoonfar** received her Ph.D. degree in Computer Science from the University of Salford, Manchester, U.K. She is currently an Associate Professor and director of Computer Vision and Remote Sensing Laboratory (Bina lab) at UMBC. Her research interests include Deep Learning, Computer Vision, Data Science, AI for Social Good, Remote Sensing, and Document Image Analysis. Her research specifically focuses on developing novel machine learning and computer vision algorithms for heterogenous sensors such as Radar, Sonar, Multi-spectral, and Optical. Her research has been funded by several awards including NSF BIGDATA award, Amazon Academic Research Award, Amazon Machine Learning award, Microsoft, and IBM.