Analysis of Aerial Images Using Deep Learning to Identify Critical Areas in Natural Disasters

Nidhya Shivakumar

The Harker School
San Jose, CA, USA
23nidhyas@gmail.com

***Abstract*—Climate change has led to a rise in the frequency and intensity of natural disasters, which, in turn, affect a wide swath of human-populated areas. The use of Unmanned Aerial Vehicles (UAVs) is on the rise in mapping out the disaster areas in their immediate aftermath. However, processing the vast amount of data obtained can take several hours to days, costing crucial time that could be used in saving lives and infrastructure. In this study, methods were developed to automate and accelerate the identification of areas that are in critical need of assistance. A Faster R-CNN object detection model was built to classify buildings into damaged and undamaged with 90% precision, and to further sub-classify them by the type of damage (undamaged, flood, rubble). The number of high quality labeled images required for training models was increased by 163% by developing an auto-label generation technique using weak supervision. Ensemble modeling further improved the recall of model predictions by 16%, and with higher prediction accuracy, when analyzing disasters not included in the training set. The utility of the model was demonstrated by using it to produce an annotated video of the 2021 tornado damage in Kentucky. A Mask R-CNN segmentation model had the best performance overall and identified undamaged roads with 100% precision and building damage with precision greater than 84% and with recall greater than 74% for all object types. This study demonstrates the power of deep learning in the processing of images from disaster-stricken areas to aid search and rescue efforts and significantly reduce disaster response times.**

***Index Terms*—Unmanned Aerial Vehicles (UAV), Faster R-Convolutional Neural Network (fr-CNN), Mask R-CNN, Weak Supervision, Ensemble Model**

I. Introduction

Over the past fifty years, climate change has led to a greater than five-fold surge in the number of global natural disasters. As reported by the World Meteorological Organization and the UN office of disaster risk reduction, in the period between 1970 and 2019, over 11,000 disasters have been reported all across the world [1] [2]. The recent Atlantic Hurricane Ian in September of 2022 is just one such example, which intensified from a category 1 to category 4 in just a period of 24 hours. Such natural disasters have resulted in over 2 million deaths and $3.64 trillion dollars in losses globally as shown in Figure 1.

Due to the resulting damage to critical power and communication infrastructure, first responders are unable to quickly and easily identify and respond to individuals in need of assistance, resulting in larger death tolls and greater financial losses. In the immediate aftermath of such disasters, it is important to garner valuable and time-sensitive information about the nature of the damage and the areas that need to be prioritized in the search and rescue efforts. To aid in relief following natural calamities, images and videos from Unmanned Aerial Vehicles (UAVs) are increasingly being used to quickly assess the damage and provide critical information to first responders. However, a key missing element in current disaster response is the ability of first responders to rapidly receive information on sections within the larger disaster zones requiring immediate assistance. Availability of this crucial piece of information will serve to significantly improve their response and enable them to prioritize their relief efforts to populations and areas with the most urgent needs. Therefore, for the most efficient response from the search and rescue team, analysis of UAV images and videos in real-time through automated classification could aid in the rapid identification of critical areas following a disaster, so that rescue efforts can be concentrated appropriately.

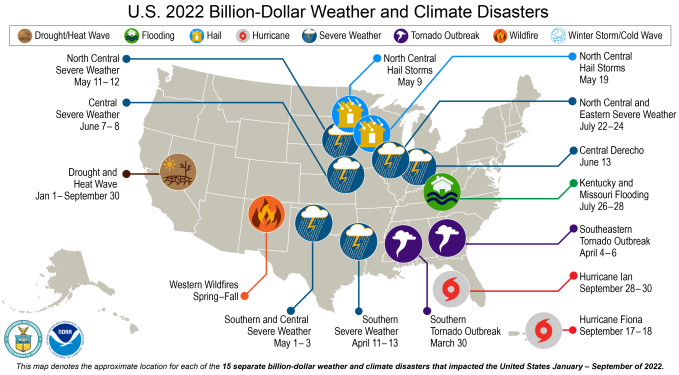

Fig. 1. Billion dollar [3] weather related disasters in the US, 2022.

A. Related Work

Combining the use of drones and artificial intelligence to aid in disaster relief efforts is still in its infancy. Early work in this area has focused on deploying drones to efficiently map an area [4] [5] [6]. Another study describes the use of a path planning algorithm and the coordination of multiple UAVs to efficiently map the disaster area [7]. These studies worked on improving data collection methodologies using UAVs to create maps of disaster sites for future offline data analysis.

The utilization of deep learning models to classify small objects in videos by following them through multiple frames has been previously achieved [8]. For example, CNN-based models have been used to detect road signs based on images and videos [9] [10], and distinguish them from background objects. Previous work has classified buildings based on terrestrial images [11]. Additionally, object detection on images obtained through synthetic aperture radar [13] and hyper spectral images [12] has enabled models to distinguish the target from background clutter.

Several earlier reports used manual assessments to analyze drone data coming in from the disaster sites, for search and rescue and for damage assessment [14] [15] [16]. However, this type of analysis is laborious and time consuming, and results were obtained days or sometimes weeks after the disaster, particularly due to the large amounts of data and the vast regions of the disaster zones.

The use of artificial intelligence and UAVs for commercial use is still nascent. Current state-of-the-art deployed by some commercial companies like Drone Sense (dronesense.com) and Cloud Factory (cloudfactory.com) focus on the complexities and logistics of deploying drones to collect data for analysis.

Over the past three years, studies on the use of AI to examine drone images and videos have emerged. One of the key use cases has been to identify damage and perform a safety assessment, including insurance appraisal [17]. In terms of the use of AI and drones in search and rescue efforts, a few reports have been published recently to identify damaged buildings in drone images [18] [19] [20]. These studies have focused on a single type of disaster; such as flood damage, or damage from earthquakes [17]. However, a key limitation of these studies is the inability to classify multiple forms of damage in the event of a disaster. For example, a hurricane could cause multiple types of damage: wind damage creating rubble and a storm surge causing flood damage. Work that is limited to studying building damage of one type may miss the multiple ways in which a building can be damaged, which is crucial for a first responder in assessing the type of relief efforts. Some studies require before and after images of the disaster site in order to accurately classify the damage status to the buildings [19] [21]. However, in most disaster scenarios, pre-disaster images may not be readily available for easy processing; and even if they are, it may be difficult to compare them to the post-disaster drone images, given the difference in how these images were acquired. The identification of roads in UAV images has been recently studied [22] [23]; however, these roads were not analyzed in the context of a natural disaster in the presence of other damages to the area.

B. Key Contributions

In this study, an automated labeling method was developed using weak supervision to automate the laborious task of manually labeling and creating bounding boxes around thousands of buildings for training models. A single frame is used to analyze the image presented by a drone without the need to rely on before and after images, which may not always be available. The highlight of this study is that the trained models are able to classify various types of damage (rubble, no damage, flood) in a given image, and is also used to predict damage to buildings from disasters that the model is not trained on: wars, earthquakes, and fires. Further, an ensemble model is created and used, which improves the prediction accuracy and recall. The model is also able to identify damaged buildings from frames of drone videos, which for example could be used in real time to identify the path of a tornado, based on the damage to buildings. In addition to identifying the buildings and classifying them in the images, a segmentation model was used to map out navigable roads. Being able to receive this complete picture of the path and destination would be extremely valuable for first responders, as it provides a realistic safety assessment of the entire disaster area and they can quickly discern the mode of transportation and route to follow to deliver aid to the critically affected areas.

II. Methods

A. Facebook Detectron2 and Google CoLab

The Detectron2 [24] library from Facebook Research was used, which provides methods to build custom object detection algorithms. Detectron2 allows users to adapt object detection and training on customized datasets with their own labels. The program was compiled through a Google Colab notebook, with Google GPU machines on the Google Cloud Platform [25]. Python code was written in several Google Colab notebooks to train and analyze multiple models and document results.

Detectron2 provides a Faster R-CNN [26] model, which enables the detection of bounding boxes around predicted objects of interest. The faster R-CNN model uses a region proposal network, which enables the prediction of object bounds and the objectness scores. All of the models were trained for 10,000 epochs and 70% of the images were used for training, and 30% were used for prediction and validation. In order to improve the prediction recall accuracy of the work, ensemble learning technique [27] was used for some of the experiments. The frCNN model was found to be inadequate in identifying road segments and in order to effectively identify navigable roads, the Mask R-CNN [28] model was implemented through the Detectron2 library to identify both buildings and roads in a single frame.

B. Low Altitude Disaster Imagery Dataset

The Low Altitude Disaster Imagery (LADI) dataset [29] was used as the base data for training the models. This dataset contained labeled images that were collected by flying a low altitude UAV over the affected areas shortly after a disaster. However, for this study, the default bounding box annotations in the LADI dataset were inadequate and therefore not usable. A custom labeling and annotation solution was developed to generate image labels for training the models.

First, any building in a given image (damaged or not) was manually labeled with bounding boxes using the LabelImg software [30]. However, labeling bounding boxes around buildings is a fairly difficult and laborious task. A weak supervision approach was used in this study to supplement the human labeled dataset, and the details are outlined in the next section.

C. Automating Image Data Labeling using Weak Supervision

Generating training data for images is very time consuming and a challenging task. In the past work, has been done to increase the accuracy of models using other approaches such as identifying Regions of Interest (ROI) in images [33]. In order to generate labeled data sets that can increase the inputs to machine learning models, weak supervision was implemented. This technique was pioneered by the Snorkel project at Stanford University [34] for models based on text data. They used models trained on a small set of manually-generated labels to produce a very large volume of high quality labeled datasets. Generating auto-labeled data using weak supervision on images is challenging, since it requires describing a group of pixels having certain characteristics [35]. This method then uses this generic description for the grouping and expands them at scale. This procedure for generating labels is done using static code analysis [36] to reduce the data required to learn the image structure. An approach based on this general principle to automate data labeling was adapted for this image-based dataset was adapted in the following way.

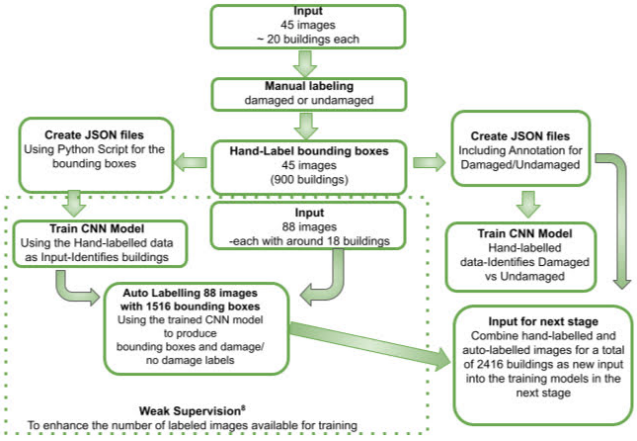

Fig. 2. Flowchart depicting auto-generation of labeled dataset using weak supervision

The preprocessing steps required for this process is shown in Figure 2. Each image taken by a drone in the LADI dataset has dozens of buildings, and manually drawing bounding boxes around these images is very time consuming. As a first step, 45 images are collected from the LADI dataset, each with an average of 20 buildings. These are manually labeled with bounding boxes around the buildings, resulting in 900 buildings being manually labeled in total. JSON [37] files are created using the labeling software. These manual labels were used to train an frCNN model, which was used to predict 1516 labeled buildings with bounding boxes from 88 images as a boost strapping mechanism to produce more training data.

The LADI images for labeling were split into two sets of images, one containing buildings without any damage and another set of images containing only damaged buildings. The models generated from the above procedure were used to draw bounding boxes around these buildings and the damage label of the overall image was assigned as the label for each individual building.

In summary, the weak supervision labeling technique was used effectively to auto label and generate 163% more training data. A combined 2416 buildings was used as input for the next step for building a frCNN model.

D. Categories of Data

In total, 2416 buildings were labeled for the training data, most of which were produced through the labeling method developed by weak supervision. 900 of these were used for validation and the rest of the data points were used for training. Table I shows the data split between training and validation. There was a spread of images from various disaster types: 37% of the buildings were undamaged, 19% were buildings that became rubble, and 42% were damaged due to floods. A similar split of buildings was generated for testing the models that were produced.

TABLE I
BUILDING DAMAGE TYPE DATA FOR TRAINING AND VALIDATION

	Num Buildings	Undamaged	Rubble	Flood
Training	1516	36%	21%	41%
Validation	900	37%	18%	43%

In addition, training data was also generated based on Google Maps to identify road segments. 25 images and 188 road segments were produced from locations in the Midwest. This was used as raw input for training models to identify road segments in images.

E. Metrics: Intersection over Union (IoU), Precision, Recall

The Intersection over Union1 metric was used for determining the accuracy of the model's prediction [31]. This metric measures how much two given boxes overlap in relation to how much area the boxes cover in total. So, if two bounding boxes each had 100 pixels and 70 pixels from each box is the intersection, the union would be 130 pixels (since the pixels at the intersection are not counted twice); so the intersection over union would be $70/130$ pixels, or around 54%.

$$\text{IoU} = \frac{\text{\# of pixels in common}}{\text{\# of total pixels}} \tag{1}$$

True Positives are the bounding boxes that the model predicts correctly as true, and True Negatives are the boxes that the model predicts as false. Both of these metrics explain how accurate the model is in predicting and classifying buildings. The Precision [32] and Recall metric derived from the True

Positive and True Negative were used as the evaluation metric for the models. Precision defines the ability of the model to accurately identify an object belonging to a category. The precision2 is the number of true positives over the total number of positives predicted.

$$\text{Precision} = \frac{\text{True Positives}}{\text{Total Positives Predicted}} \tag{2}$$

The recall metric was used to evaluate the number of objects that met the criteria for detection but not accurately identified as an object meeting that criteria. The recall3 is the number of true positives over the number of actual positives.

$$\text{Recall} = \frac{\text{True Positives}}{\text{Actual Number of Positives}} \tag{3}$$

These two metrics together provide a good evaluation of the effectiveness of the object detection model. For the experiments, the models are tuned for very high precision, with modest recall greater than 70%.

III. Results and Discussion

A. Building Damage Identification using frCNN Model

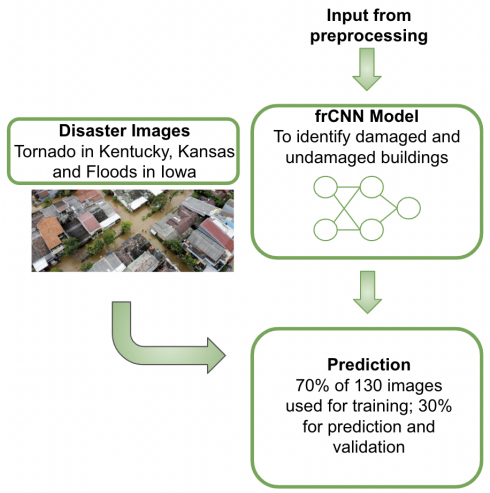

Fig. 3. Workflow for Building Damage Identification

The images and bounding box annotations with labels were used as an input for this step. The process for this step is outlined in Figure 3. A Faster R-CNN model was trained to identify damaged and undamaged buildings in an image. In addition to using the trained model to predict bounding boxes around buildings in images from the LADI dataset, buildings in disaster images from tornadoes and floods in Kentucky, Kansas, and Iowa were predicted as well.

Figure 4 shows the images the model produced after training. The left-hand side shows the raw images, and the right-hand side shows the model's predictions on these images. These are images from the floods in Iowa and tornadoes in Kentucky. In these images, 0 refers to an undamaged building and 1 means damaged building[1]. Images in Figure 4a and Figure 4c show flood damage to buildings, and Figure 4b and 4d show the model accurately predicting that the buildings marked by the bounding boxes were affected by flood. Figure 4e shows a few buildings that were damaged by the tornadoes in Kentucky, and the image in Figure 4f shows the labels produced by the model. In this particular example, the model missed marking one of the buildings as damaged.

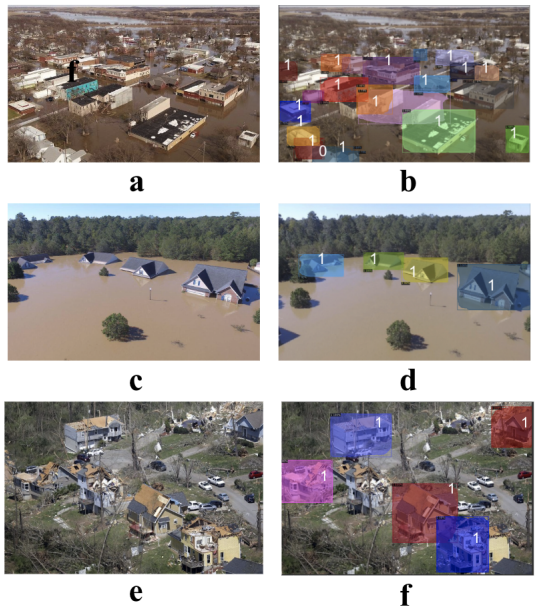

Fig. 4. Images of 2021 Kentucky floods. A label of 0 means undamaged, and 1 means damaged. The images on the left-hand side show the raw images, and the right-hand side shows the images after the predictions.

TABLE II
Precision and Recall values for Damaged and Undamaged Buildings

	Avg Recall	Avg Precision
Undamaged	89.66	97.50
Damaged	66.12	94.13

Overall the model performed well and predicted the bounding boxes for most undamaged buildings, and had a very high precision of over 97% and a recall of close to 90% as shown in Table II. However, for detecting damaged buildings, although the prediction precision was high, the recall was closer to 66%. This result is expected since the model is tuned to deliver high precision results. This is largely attributable to the fact that damage comes in different forms and increasing the recall to identify damage requires more image data for training.

[1]Note: The 0's and 1's printed by the software on the images were not clear so these 0's and 1's were manually added to the exact same boxes as the model to make it easier to see what the model predicted

B. Building Identification and Damage Categorization Using frCNN Model

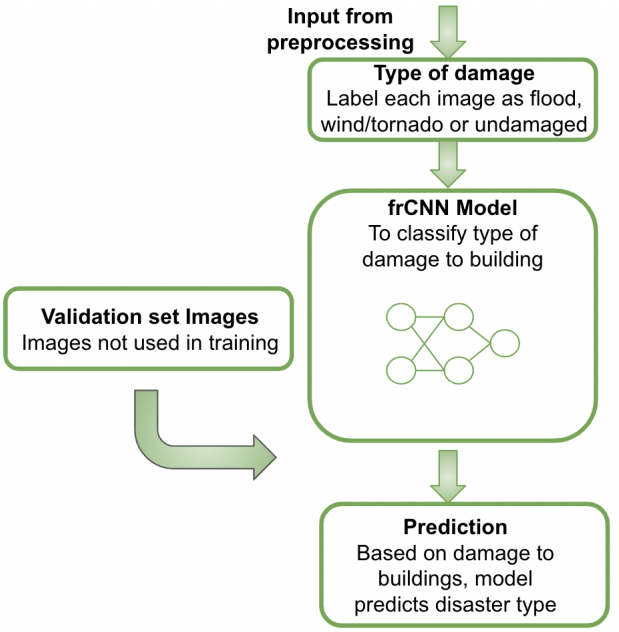

Fig. 5. Flowchart for Damage Categorization

In this experiment, a model was trained to go beyond binary classification of damaged or undamaged, and identify the type of damage to a building. For example, Hurricane Ian recently ravaged various parts of Florida, and it caused both flood and wind damage to buildings. As a result, there is a necessity to classify the buildings into types of damage to provide appropriate assistance. The labeled annotations for the images used in training were modified for this experiment and the pre-processing steps were implemented using weak supervision to generate labeled annotations: undamaged, rubble(representing damage other than flood), and flood. These were used as input into the next step for model training. The process workflow described above was used to generate the labels and passed into training the models as shown in Figure 5.

This data was used to train an frCNN model, which was able to predict the bounding boxes around the buildings and classified these buildings with the appropriate damage types.

The resulting images are shown in Figure 6. Figure 6a shows the image of a neighborhood with no damaged buildings and Figure 6b shows the bounding boxes and labels predicted by the model. All of the buildings in the image in Figure 6b were accurately identified with label 0, or undamaged. Figure 6c and Figure 6d are images from a flood in Kentucky, and most of these buildings were accurately labeled 2, or flood damaged. There are two buildings on the bottom right of the image that were labeled as 1 or rubble, because in some cases the model was not adequately able to resolve whether the building was damaged due to floods. Figure 6e shows an image of destruction caused by a tornado; however, this tornado only damaged the four buildings on the left of the image, leaving

Fig. 6. Tornadoes and Floods in the Midwest in 2021. A label of 0 means undamaged, 1 means rubble, and 2 means flood damage.

the other two undamaged. The model was able to identify this difference, as shown by the labels in Figure 6f, labeling the buildings with some damage as rubble shown as 1, and the two undamaged buildings on the right as 0.

TABLE III
PRECISION AND RECALL VALUES FOR UNDAMAGED, RUBBLE, AND FLOOD DAMAGED BUILDINGS.

	Avg Precision	Avg Recall
Undamaged	100	89.66
Rubble	97.65	69.17
Flood	88.37	60.32

Table III shows the average precision and recall for the various types of damage to buildings being predicted: undamaged, rubble, and flood. Similar to earlier observations, the average precision of the model prediction is very high at close to 90%. The recall for identifying buildings is once again fairly high at close to 90%. The recall for classifying the buildings damaged by floods and rubble is 60% and 69% respectively. The recall numbers are strong for a generalized object detection model that are tuned to such a high precision. These models are tuned to deliver very high precision results even at the expense of recall so that first responders would have an accurate estimate of what to expect in a disaster scene.

C. Classifying Images from Untrained Disasters

The images and modified annotations from the previous experiment were also used as input to train the models for this experiment. In this experiment, the model was evaluated

on images from three disasters previously not trained on: earthquakes, war, and fire.

When the model was applied on images from these disasters, the precision continued to be around 90% but the recall dropped down to 52% from 65%. One of the key reasons for this drop is the fact that although these images represented damaged buildings, the appearance of the buildings post-disaster is different from the images with which the models were trained on, especially for images based on disasters that the model was never trained on. In order to overcome this drop in recall, ensemble modeling was employed to boost the prediction accuracy of the ML model and this is detailed in the next sub-section.

D. Improving Model Recall using Ensemble Model

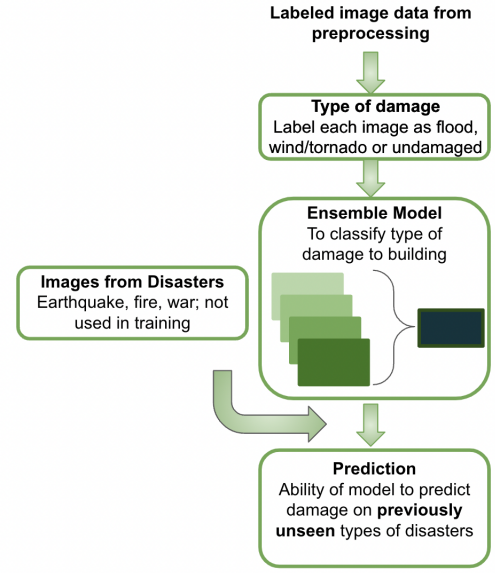

Fig. 7. Workflow for Ensemble Model

Ensemble learning [27] involves combining the many individually trained models, resulting in higher performance when compared to each of the individual models. The logic behind the ensemble is that neural networks start out with a set of random weights, and depending on the starting weights, it could converge to different values, which would result in different predictions for the same image. By combining multiple models with a different set of starting weights and varying predictions, the maximum prediction accuracy can be achieved.

Figure 7 shows the flowchart for this process. Four different models were trained on the input. With the addition of each of these models to the ensemble, the recall of the combined model improved, as shown in Table IV. Figure 8 gives an intuition behind how the ensemble model works. Visually, the output of the ensemble model predicted more buildings and their correct damage status when compared to the individual models. The ensemble model as depicted in Figure 8 was able to predict more buildings in images labeled as being damaged, improving the recall.

TABLE IV
RECALL OF ENSEMBLE MODEL. THIS TABLE SHOWS THE RECALL OF THE ENSEMBLE MODEL BASED ON THE NUMBER OF MODELS THAT ARE AGGREGATED. AS THE NUMBER OF MODELS INCREASES, THE RECALL IMPROVES.

# of Models	Recall %
1	52.5
2	56.1
3	58.9
4	60.9

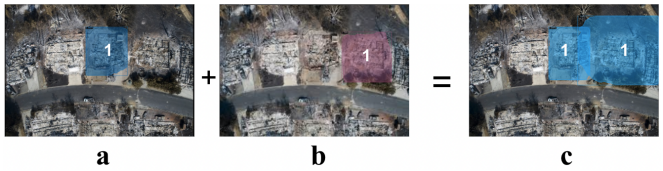

Fig. 8. Results of Ensemble Model. Images a and b each have one box labeled, but when combined using the ensemble model, there are two boxes labeled in the image to the right which produces a super set of these two images.

The images that this ensemble predicted had many more buildings classified accurately than the models individually. The results are shown in Figure 9.

E. Frame-by-Frame Analysis on Drone Videos

In the event of a disaster, drones are capable of providing an aerial vantage point of the disaster areas for the first responders. An annotated video detailing the areas of damage would be extremely valuable for first responders to be able to direct the accurate aid to the appropriate location. A drone video from the town of Bowling Green, Kentucky [38] was obtained, shortly after the tornado in December of 2021. Additionally, a drone was flown around a neighborhood, in order to generate a video of undamaged homes. Cv2 [39] was used to generate image frames from both of these videos, and the model was used to predict on these images, as outlined in Figure 10. After the bounding boxes and labels were predicted, the frames of each video were stitched back together, producing an annotated video. This video that was generated, demonstrating the viability of this method, is available for viewing and the details are in Shivakumar [40].

A few frames from the video are shown in Figure 11, and the path that the tornado followed through the town is clearly visible. The buildings toward the center of the frame are labeled as rubble (red box), but the ones on the outside are normal (yellow box). This type of annotated video depicting the damage would be very useful to first responders in the immediate aftermath of a disaster.

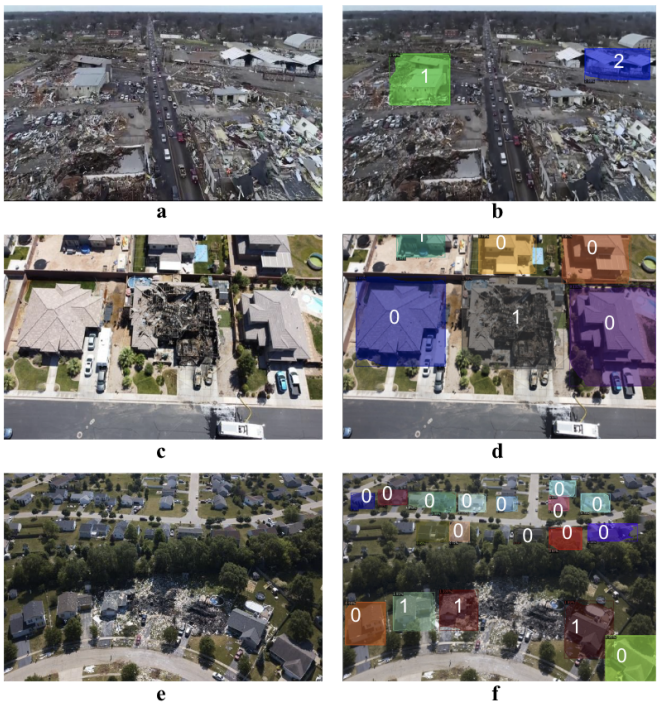

Fig. 9. Ensembling on Previously Unseen Disaster Types. Images from disasters such as earthquakes, fires, and even war, not used in training, were analyzed using the ensemble model. The images on the left are original, and the ones on the right are the predictions by the ensemble model.

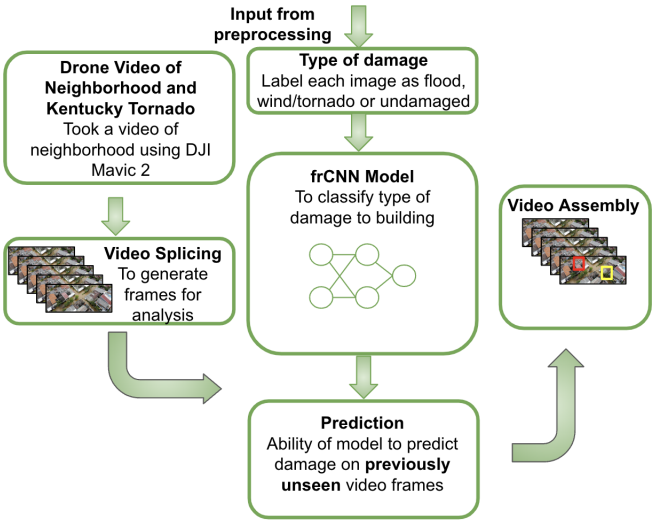

Fig. 10. Analysis of Drone Video. The frames from the video were spliced, and each frame was analyzed by the model. Following the predictions, the frames were stitched back together, creating the annotated video.

Fig. 11. Model predictions on select frames of the video. These images are from the tornado disasters in Kentucky. In both of these images, the path that the tornado took through this town is clear, because the buildings in the center are labeled as damaged, whereas the ones on the outside are undamaged.

F. Building Classification and Roads Segmentation using Mask R-CNN

Identifying roads that are navigable is an important component of enabling faster disaster relief. The frCNN modeling was first extended to be able to train on and identify roads that are undamaged. The trained model draws bounding boxes around identified objects. While bounding boxes work well to identify buildings, this is difficult to use for roads since their shapes are not well defined. In order to adequately model and identify roads a switch to using the segmentation method [41] was made, that uses the Mask R-CNN [28] model. The segmentation works differently, predicting whether the given pixel is a part of a road or not for each pixel in the image.

In order to label objects for segmentation modeling, the label imaging software [30] was used, in which four points were used to encompass the road, instead of a box. These labeled images were used as input into a machine learning model. The initial training of the model was only for the purpose of identifying road segments.

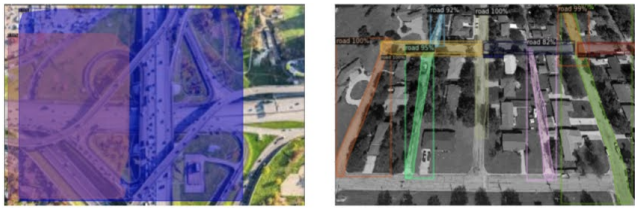

Fig. 12. Faster R-CNN vs Mask R-CNN for road predictions. The figure on the left shows the regular Faster R-CNN predictions for roads. Since the predictions are shown as rectangles, it is difficult to accurately outline the road. The image on the right shows the Mask R-CNN predictions. This model predicts whether each pixel in the image is a road or not, and therefore, outlines the road clearly.

Figure 12 shows a comparison of the frCNN model prediction for roads (Figure 12a) and the segmentation model (Figure 12b). As seen in Figure 12a, the bounding box encompasses a large portion of the image, and does not clearly outline the road. However, in Figure 12b, the exact shape of the road is outlined with segmentation. Just to note, this particular model was purely detecting roads in an image, without buildings (a combined road and building model was trained later).

Figure 13 details the process used to evaluate the models with roads. The new label annotations for the roads using

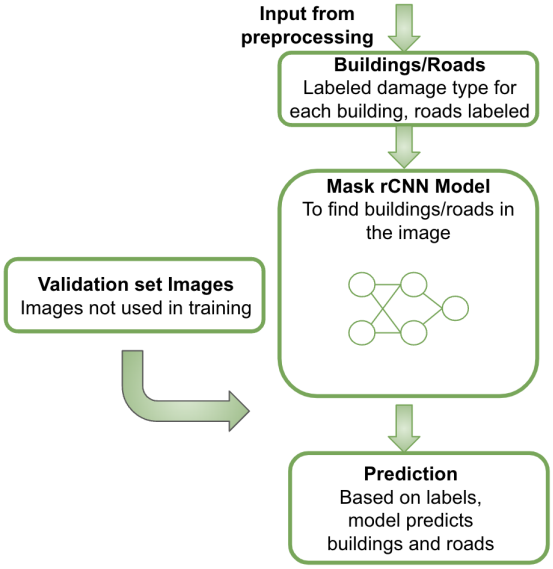

Fig. 13. Workflow for Mask R-CNN segmentation model for road predictions

segmentation along with the previous labels for building annotations were used as input to train a combined model to identify both roads and buildings. Figure 14 shows the results of the segmentation performed on roads and buildings combined. This type of analysis depicts the clear outline of the road as discussed earlier, along with outlines of damaged or undamaged buildings.

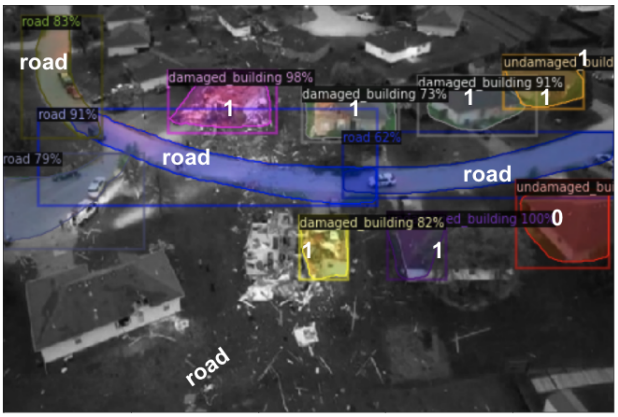

Fig. 14. Simultaneous Road and Building predictions. The image shows the segmentation of the road along with the classification of the building damage. Some of the roads in the image are not labeled, which is due to the fact that they were too damaged for the model to identify it as a road at all.

Table V shows the precision and recall of the model on the validation data set. The precision was greater than 84% for this model in all the damage categories, with recall above 74%, which was significantly better than the frCNN model. In addition, the precision was 100% for detecting roads and the recall was greater than 77%. The results of this experiment clearly demonstrate that the Mask R-CNN is far superior to the frCNN model in being able to identify buildings and roads in the same image.

TABLE V
PRECISION AND RECALL SCORES FOR BUILDINGS, ROADS, DAMAGED, AND FLOODED BUILDINGS AND VALIDATION SET.

	Avg Precision	Avg Recall
Building	100	77.55
Road	100	74.29
Damage	96.97	94.12
Flood	84.63	77.78

IV. Conclusions and Future Work

Through this research, accurate prediction of building damage from images of recent disasters was achieved, along with an extension of this prediction to previously unseen disasters, providing proof on the extensibility of this model. The model in this study was effective in classifying different types of damage in a given frame, information which is crucial for first responders as multiple damage types can exist in a single disaster zone, and the rescue efforts will be dependent on the type of damage. The demonstration of the effectiveness of this type of analysis in video footage and the generation of a frame-by frame analyzed video provides a vision of how the data will be made available to first responders. Since the analysis and predictions will be performed in real-time, a very rapid response time is feasible.

In general, the most tedious and time-consuming aspect of building ML models with image data is the generation of labels. The power of this work stems from the ability to scale the manual labeling of images through automation by using weak supervision. The Faster R-CNN model developed in this study was able to accurately, and with high precision and recall, classify multiple types of damage to buildings. The superiority of the Mask R-CNN segmentation model was evident from its improved prediction accuracy over the Faster-R-CNN model and the identification of navigable road segments in disaster images with high precision. Using an ensemble modeling approach, the study extended the prediction of damage to images from disasters that the model was not trained on. The development and successful use of techniques such as ensemble modeling and weak supervision demonstrate that this analysis can be extended to other image processing applications as well.

In future work, the images produced by the drones will be mapped to GPS coordinates to precisely provide navigation instructions for first responders to quickly reach their destination. The drone itself has data about its location, and this, along with the angle of the camera, can be used to calculate the GPS coordinates. Further, the panoramic stitching of object and damage predictions from multiple images from one or more UAVs can provide a more realistic and multidimensional perspective of the damage assessment to buildings and infrastructure.

Certain types of disasters may cause damage to infrastructure, and if a drone is sending videos and photos through the cell tower, that may not be possible due to significant damage.

If this connectivity is limited due to the damage, the first responders will experience difficulty in receiving the images from the drones in real-time. Therefore, a future extension of this project which is in the planning stage is to have much of the intelligence live in the drone itself and for the drone to only send the analyzed results, which will require limited connectivity.

Another limitation of the model is the low recall for previously unseen types of damage that the model has not trained with. While ensembling is a powerful technique, training the model with more labeled data can further improve the performance significantly. In addition, the creation of a large database of disaster site images in the future from multiple sources that will incorporate satellite imagery and cell phone data will extend the effectiveness of this study.

REFERENCES

[1] International Federation of Red Cross and Red Crescent Societies "World Disasters Report 2020: Come Heat or High Water - Tackling the Humanitarian Impacts of the Climate Crisis Together [EN/AR]," https://reliefweb.int/report/world/world-disasters-report-2020-come-heat-or-high-water-tackling-humanitarian-impacts, September 17, 2021.

[2] World Meteorological Organization, "Weather-related disasters increase over past 50 years, causing more damage but fewer deaths," https://public.wmo.int/en/media/press-release/weather-related-disasters-increase-over-past-50-years-causing-more-damage-fewer, Aug31, 2021

[3] NOAA National Centers for Environmental Information (NCEI) U.S, "Billion-Dollar Weather and Climate Disasters," *https://www.ncei.noaa.gov/access/billions/,* DOI: 10.25921/stkw-7w73, 2022

[4] S. Adams, and C. Friedland, "A Survey of Unmanned Aerial Vehicle (UAV) Usage for Imagery Collection in Disaster Research and Management," *Research Gate* 2011.

[5] T. Zweglinsky, "The Use of Drones in Disaster Aerial Needs Reconnaissance and Damage Assessment - Three-Dimensional Modeling and Orthophoto Map Study," *Sustainability 2020,* 12, 6080. https://doi.org/10.3390/su12156080.

[6] F. Greenwood, E. Nelson, and G.P. Greenough, "Flying into the Hurricane: A Case Study of UAV Use in Damage Assessment during the 2017 Hurricanes in Texas and Florida," *PLOS ONE, vol. 15, no. 2,* Feb. 2020, p. e0227808. doi:10.1371/journal.pone.0227808.

[7] O. Artemenko, A.Rubina, O.Golokolenko, T.Simon, J.Romisch, A.Mitschele-Thiel, "Validation and evaluation of the chosen path planning algorithm for localization of nodes using an unmanned aerial vehicle in disaster scenarios," *6th International Conference on Ad Hoc Networks,* DOI:10.1007/978-3-319-13329-4_17, Rhodes, Greece, August 2014.

[8] Y. Liu, L. Geng, W. Zhang, Y. Gong, and Z. Xu, "Survey of Video Based Small Target Detection," *Journal of Image and Graphics,* Vol. 9, No. 4, pp. 122-134, December 2021. doi: 10.18178/joig.9.4.122-134

[9] F. Utaminingrum, R. P. Prasetya, and Rizdania, "Combining Multiple Feature for Robust Traffic Sign Detection," *Journal of Image and Graphics,* Vol. 8, No. 2, pp. 53-58, June 2020. doi: 10.18178/joig.8.2.53-58

[10] R. Hasegawa, Y. Iwamoto, and Y-W. Chen, "Robust Japanese Road Sign Detection and Recognition in Complex Scenes Using Convolutional Neural Networks," *Journal of Image and Graphics,* Vol. 8, No. 3, pp. 59-66, September 2020. doi:10.18178/joig.8.3.59-66

[11] T. Chaloeivoot and S. Phiphobmongkol, "Building Detection from Terrestrial Images," *Journal of Image and Graphics,* Vol. 4, No. 1, pp. 46-50, June 2016. doi: 10.18178/joig.4.1.46-50

[12] E. Lo, "Target Detection Algorithms in Hyperspectral Imaging Based on Discriminant Analysis," *Journal of Image and Graphics,* Vol. 7, No. 4, pp. 140-144, December 2019. doi: 10.18178/joig.7.4.140-144

[13] S. Huang, Y. Wang, and P. Su, "A New Synthetical Method of Feature Enhancement and Detection for SAR Image Targets," *Journal of Image and Graphics,* Vol. 4, No. 2, pp. 73-77, December 2016. doi: 10.18178/joig.4.2.73-77

[14] D. Câmara, "Cavalry to the rescue: Drones fleet to help rescuers operations over disasters scenarios," *2014 IEEE Conference on Antenna Measurements and Applications (CAMA),* 2014, pp. 1-4, doi: 10.1109/CAMA.2014.7003421.

[15] P. Tatham, "An investigation into the suitability of the use of unmanned aerial vehicle systems (UAVS) to support the initial needs assessment process in rapid onset humanitarian disasters," *2009, International Journal of Risk Assessment and Management,* v. 13, n. 1, p. 60-78

[16] C.A.F. Ezequiel, M. Cua, N.C. Libatique, G.L. Tangonan, R. Alampay, R.T. Labuguen, C.M. Favila, J.L.E. Honrado, V. Canos, C. Devaney, C, A.B. Loreto, J. Bacusmo, J. B. Palma, "UAV aerial imaging applications for post-disaster assessment, environmental management and infrastructure development. 2014 International Conference on Unmanned Aircraft Systems," *ICUAS 2014 - Conference Proceedings,* n. 6842266, p. 274-283

[17] A. Calantropio, and F. Chiabrando, M. Codastefano, and Bourke, E, "Deep Learning For Automatic Building Damage Assessment: Application in Post-Disaster Scenarios using UAV Data," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* V-1-2021. 113-120. 10.5194/isprs-annals-V-1-2021-113-2021.

[18] S. Naili, M. Inthizami, M. Anwar, A.R. Machmud, G. Ahmad, A. Ronni, Kurnianingsih and J. Wisnu, "Flood video segmentation on remotely sensed UAV using improved Efficient Neural Network," *ICT Express,* Volume 8, Issue 3, 2022, Pages 347-351, https://doi.org/10.1016/j.icte.2022.01.016

[19] Y. Pi, N.D. Nath and A.H. Behzadan, "Convolutional neural networks for object detection in aerial imagery for disaster response and recovery," *Advanced Engineering Informatics,* Volume 43, 2020, https://doi.org/10.1016/j.aei.2019.101009.

[20] C. Kyrkou and T. Theocharides, "Deep-Learning-Based Aerial Image Classification for Emergency Response Applications Using Unmanned Aerial Vehicles," *arXiv.1906.08716,* June 2019, https://doi.org/10.48550/arXiv.1906.08716.

[21] H.S. Munawar, F. Ullah, S. Qayyum, S.I. Khan, M. Mojtahedi, "UAVs in Disaster Management: Application of Integrated Aerial Imagery and Convolutional Neural Network for Flood Detection," *Sustainability 2021,* 13, 7547. https://doi.org/10.3390/su13147547

[22] M. G. Bocanegra and R. J. Haddad, "Convolutional Neural Network-Based Disaster Assessment Using Unmanned Aerial Vehicles," *SoutheastCon 2021,* 2021, pp. 1-6, doi: 10.1109/SoutheastCon45413.2021.9401928.

[23] V. Pham, C. Pham and T. Dang, "Road Damage Detection and Classification with Detectron2 and Faster R-CNN," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 5592-5601, doi: 10.1109/BigData50022.2020.9378027

[24] Y. Wu, A. Kirillov, F. Massa, W-Y. Lo and R. Girshick, "Detectron2", https://github.com/facebookresearch/detectron2,2019

[25] E. Bisong, "Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginner", 59–64 (Apress, 2019).

[26] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015, https://github.com/ShaoqingRen/faster_rcnn.

[27] M.A. Ganaie, Minghui Hu, A.K. Malik, M. Tanveer, P.N. Suganthan, "Ensemble deep learning: A review," *Engineering Applications of Artificial Intelligence,* Volume 115, 2022, 105151, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2022.105151

[28] W. Abdulla, "Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow," *Github Repository, https://github.com/matterport/Mask_RCNN*, 2017,

[29] J. Liu, D. Strohschein, S. Samsi and A. Weinert, "Large Scale Organization and Inference of an Imagery Dataset for Public Safety," *2019 IEEE High Performance Extreme Computing Conference (HPEC),* Waltham, MA, USA, 2019, pp. 1-6. doi: 10.1109 HPEC.2019.8916437. https://github.com/LADI-Dataset/ladi-overview

[30] Tzutalin, "LabelImg Git code," 2015, https://github.com/tzutalin/labelImg

[31] P. Azevido, "Object Detection State of the Art 2022," *https://medium.com/@pedroazevedo6/object-detection-state-of-the-art-2022-ad750e0f6003*

[32] J. Hui, "mAP (mean Average Precision) for Object Detection," *Posted March 6, 2018 on Medium,* https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173

[33] R. Khan, T. Fariha Raisa, and R. Debnath, "An Efficient Contour Based Fine-Grained Algorithm for Multi Category Object Detection," *Journal of Image and Graphics,* Vol. 6, No. 2, pp. 127-136, December 2018. doi: 10.18178/joig.6.2.127-136

[34] A. Ratner, S. Bach, H. Ehrenberg, J. Fries, S. Wu, and C. Re, "Snorkel: Rapid Training Data Creation with Weak Supervision," *2017, VLDB Endowment,* vol 11, no 3, issn 2150-8097, https://doi.org/10.14778/3157794.3157797

[35] P. Varma, B. He and C.Re, "Snorkel for Image Data," , *https://www.snorkel.org/blog/coral*

[36] P. Varma, B. He, P. Bajaj, I. Banerjee, N. Khandwala, D.L. Rubin, and C. Ré, "Inferring Generative Model Structure with Static Analysis," https://arxiv.org/abs/1709.02477

[37] F. Pezoa, J.L. Reutter, F. Suarez, M. Ugarte, and D. Vrgoc, "Foundations of JSON schema," *In Proceedings of the 25th International Conference on World Wide Web*, pp. 263–273, 2016,

[38] whas11.com, "Total devastation, Drone video shows Bowling Green, Kentucky tornado damage", *https://www.whas11.com/article/news/kentucky/bowling-green-total-devastation-kentucky-tornado-damage-drone-video/417-9184c721-7a4b-4b99-95a3-6e378f0fc178*

[39] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools,* 2008-01-15 19:21:54, 2000

[40] S. Nidhya, "Video of Real-Time Analysis of Aerial Images to Identify Critical Areas in Natural Disasters," *h*ttps://www.youtube.com/watch?v=NbGG1ezXUJo&t=6s

[41] C. Andy and A. Chaitanya, "Going beyond the bounding box with semantic segmentation," *The Gradient, 2018,* https://thegradient.pub/semantic-segmentation