A Novel Disaster Image Data-set and Characteristics Analysis using Attention Model

Fahim Faisal Niloy, Arif, Abu Bakar Siddik Nayem, Anis Sarker, Ovi Paul, M. Ashraful Amin,
Amin Ahsan Ali, Moinul Islam Zaber, AKM Mahbubur Rahman.
AGenCy Lab, CSE, Independent University, Bangladesh; CSE, Dhaka University, Bangladesh.
niloy9542@gmail.com, [1611041, 1510190, 1521745, 1531144, aminmdashraful, aminali]@iub.edu.bd, zabermi@gmail.com, akmmrahman@iub.edu.bd

Introduction

Disasters have increasingly become very prevalent throughout the world due to climate change, world-wide temperature rise and many other reasons. Therefore, in the occassion of any disaster event, emergency response is crucial to save human lives and prevent losses. Automated systems can help in this regard. An effective classification algorithm that can detect disaster events in real time from images circulating in social media sites or online news portals and subsequently report to the corresponding rescue team is of great importance. But to develop such an algoirthm using deep learning, a diversified and organized disaster dataset is a must. Surprisingly, such datasets are very scarce

Motivation

Existing disaster datasets suffer from major limitations, such as:

- Low diversity
- Few classes
- Narrow regions
- No subcategories
- Dataset bias

A major limitation of existing datasets is poor attention localization capability.

Figure 1: Poor attention localization

Therefore, our motivation is to compile a comprehensive disaster dataset that doesn't suffer from any of the limitations described above. In short, we tried to compile a dataset that is much diverse, which has a lot of challenging image. Also we didn't focus only on a couple of regions, rather we tried to gather disaster images from all major regions around the world. Our dataset is the first disaster dataset that has organized sub categories. This will help researchers to detect disaster classes more specifically. Also, as a result of having diverse images, attention localization capability of classifiers trained on our dataset is much improved.

Figure 2: Sample Images from our proposed dataset

Dataset Collection and Annotation

To create our dataset we have collected images from social media sites, online news portals and available public datasets. We used hashtags with prominent disaster names to gather disaster specific images. For images in non-disaster category we mostly used public datasets made available by other researchers.

After collecting a wide pool of images we first discarded images those were very low resolution and had other problems. Rest of the images went into the annotation process. Three annotators separately annotated each image to a disaster category. After the annotation process each image was kept to the category determined by majority votes. Rest of the images were discarded. We also annotated bounding boxes to 200 images. Six annotators put bounding boxes to the regions where there attentions were intuitively drawn to infer the disaster category. The common area of all six annotators are later taken.

Damaged Infrastructure Fire Disaster

Figure 3: Bounding-box Annotations

Diversity Analysis

Here we have calculated the lossless JPG size of average image for each class. The more diverse the class, the lesser the average image size. It is observed that four out of six classes of our dataset has lesser average image size than UCI dataset. So, our dataset is more diverse than UCI.

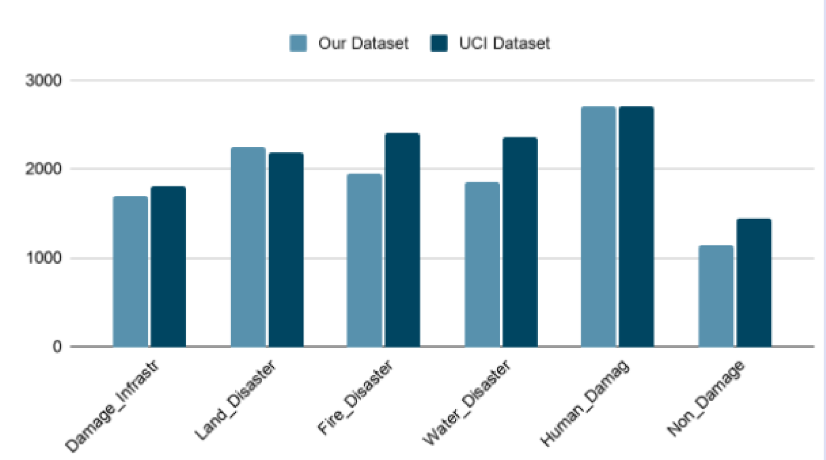

Figure 4: Lossless JPG size in byte of our dataset vs UCI

Cross Validation

The cross validation is performed in five folds. Both Class Acitavtion Map(CAM) and Three Layer Attention Map(TLAM) architectures are used. The accuracy in each fold doesn't deviate much, which shows the uniformity of our dataset.

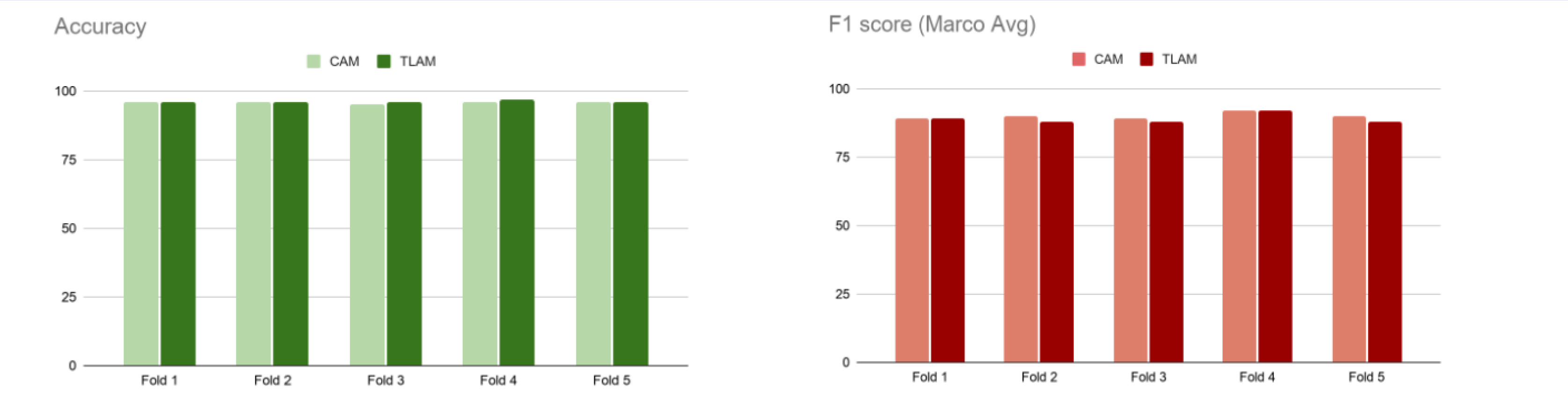

Figure 5: Cross Validation Summary for CAM and TLAM

Attention Agreement and Visualization

In attention agreement experiment we have compared binarized attention heatmap with annotated bounding boxes to calculate the mean IOU. We did the experiment with both our dataset and UCI dataset. We used both CAM and TLAM methods to generate the heatmaps. Our dataset yields better mIOU than UCI dataset in both the methods.

Table 1: Overlap agreement between human attention and classifier attention

	Human-Classifier Attention Overlap (Mean IoU)	Human-Classifier Attention Overlap (Mean IoU)
Training Set	Proposed Dataset	UCI Dataset
CAM	0.53	0.45
TLAM	0.31	0.18

Below are some examples of attention verification. The top row is the input images and middle is output from the model that was trained with UCI dataset. The last row is output from the model that was trainied with our proposed dataset. In the middle row dataset bias is clearly observed. Classifier's attention is placed in wrong regions. But in the last row, correct attention localization is observed. It has been made possible for the high diversity of our proposed dataset.

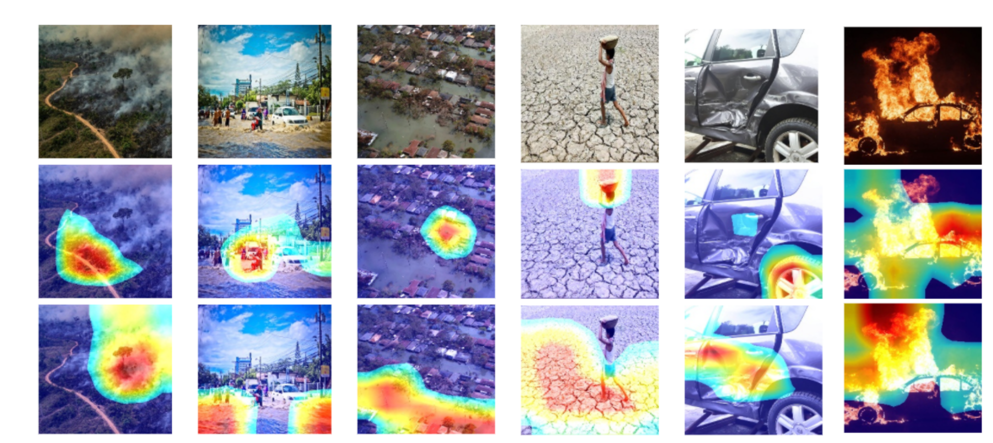

Figure 6: Top row: Input images; Middle row: Misplaced attention (Using CAM trained with UCI dataset); Last row: Correct attention (Using CAM trained with our dataset)

Conculsion

As a future work, we will annotate each image in a multi class manner, if more than one disaster event is observed in the image. Also we want to annotate bounding boxes for all the images. We hope our dataset helps the deep learning community to pursue impactful research on disaster classification.