Application of Very Deep Neural Nets to Videos Depicting Extreme Events

Team COVIS: Yuan Li*, Haili Wang*, Shuo Sun*, and Bill Buckles†
Department of Computer Science and Engineering
University of North Texas, USA
*{yuanli4, hailiwang, shuosun2}@my.unt.edu, †{bill.buckles}@unt.edu

It is well-known that the trend of increasing global natural disasters of natural disasters globally, accompanied by increasing loss of life and property, shows no signs of halting. The fifth assessment report of the Intergovernmental Panel on Climate Change (IPCC, 2014) predicts that as global warming continues in the coming decades, its contribution to the increase in natural disaster losses will become more prominent. However, through rapid and accurate analysis of disaster scenarios, there is still an opportunity to significantly reduce catastrophic losses caused by extreme events. From a video, we extract key frames and identify embedded objects (using YOLOv3).

Research will center on the Low Altitude Disaster Imagery (LADI) [1] datasets, with emphasis on disaster response applications, assessing data collected at low altitudes, and utilizing enhanced visual recognition.

The test data set – LADI dataset contain 41 original full videos and 1,825 segmented short video clips between 2 to 20 seconds. First, we prepare the segmented video clips and extract key-frames that contain the information in multiple scenes. Next, we manually filter out the minimum number of frames that can represent a specific scenario. With the data cleaning and pre-processing stages complete, we then utilize YOLOv3 for foreground object detection and filter out useful images to the experimental data.

During the key frames extraction process, we use local maxima as the final step. From a video, the inter-frame differences are computed. Using local maximum, the frames for which the average inter-frame difference are local maxima are selected as key frames. The extraction results obtained via this method perform better in diversity, and the extraction results are evenly dispersed within the video.

Considering that YOLOv3 has been widely used in practice, we will develop a new object detector based on the characteristics of YOLOv3. We will attempt to combine various existing artifices that will not increase the number of model parameters and FLOPs, to achieve the goal of improving the accuracy of the detector while ensuring that the speed is almost unchanged. As for other neural network products, we will explore different ones as backbone networks. Using the same data enhancement method as yolov4. We will analyze the factors causing misidentification including that caused by the bias of the data itself. Optimization of the existing YOLO model will be completed at two levels: The first level being improvement of the detection speed; The second level is the improvement of the performance (speed and accuracy) of the experimental results.

Aspects within the first level that are to be addressed view the yolov3 as a one-stage anchor-based detector. The detector comprises a backbone network, a detection neck (typically an FPN), and a detection head for object classification and localization. We will replace the original backbone, DarkNet-53, with ResNet50 for a one-stage anchor-free detectors based on anchor-point. We hypothesize the performance of anchor-free detectors can also compete with state-of-the-art anchor-based detectors. More importantly, we will consider using a combination of center-based and corner-based methods to improve the detection speed.

Secondly, the last level would encompass improvement of the performance of the results and reducing the impact on the entire network, we will generally modify the loss that affects only the training process. We will add IoU Loss (keep L1-loss and add branch), IoU Aware (measure the accuracy of localization) and Grid Sensitive (YOLOv4 method) modules.

Accuracy of the objects reviewed and identified will enhance the overall improvement in classifying methods within a disaster category. Finally, the description of the degree of disaster damage is determined by the combination of disaster categories and key objects with response weights.

References

[1] G. Awad, A. A. Butt, K. Curtis, Y. Lee, J. Fiscus, A. Godil, A. Delgado, J. Zhang, E. Godard, L. Diduch, J. Liu, A. F. Smeaton, Y. Graham, G. J. F. Jones, W. Kraaij, and G. Quénot. Trecvid 2020: comprehensive campaign for evaluating video retrieval tasks across multiple application domains. In *Proceedings of TRECVID 2020*. NIST, USA, 2020.