Deep Learning based Autonomous Accident Detection and Assessment†

Michael Stepzinski
University of Nevada, Reno
mstepzinski@unr.edu

Mei Yang
University of Nevada, Las Vegas
mei.yang@unlv.edu

Vehicle accidents occur frequently, which may cause severe traffic congestion and delays. The compounding factors for delays caused by accidents are accident severity, authority response time, and accident scene investigation time. The manual assessment of accident scenes can take hours with significant manpower and may involve human errors. Autonomously assessing accident scenes can speed up the scene assessment process and improve accuracy of scene analysis for transportation authorities. The goal of this research project is to develop novel drone-based methods for autonomous accident detection and assessment. This paper is focused on investigating efficient computer vision models to detect and assess accident scenes as well as evaluating the performance and computational cost of these models.

In the literature, many machine learning methods were proposed to perform similar tasks. Classifiers including Support Vector Machines, Logistic Regression, Decision Trees, Random Forests, AdaBoost, and Gradient Boosting were used to classify accident severity. Convolutional Neural Network (CNN) and You Only Look Once (YOLO) based methods were proposed for detecting accident vehicles or damaged roadside objects but lacked detection of other features. Video-based accident detection models using Resnet50, Faster R-CNN, and LSTM did not detect fires or smoke. The target application needs a light-weight deep learning based model that uses object localization to accurately detect accident features from images or videos. In addition, there are no large and labeled accident image datasets publicly available.

In this work, a computer vision model based on YOLOv5 is generated to detect accident features and an accident image dataset is compiled and labeled. YOLOv5 was chosen as it is the latest YOLO model that allows for accurate real-time localization and classification of many desired accident features in images and video. To detect accident features including regular vehicles, damaged vehicles, fire, smoke, and pedestrians, a dataset was created using images gathered from Las Vegas freeways and accident & disaster datasets. Every instance of these features was labeled in each image gathered. The original raw dataset consists of 964 images. These images were augmented to 2,314 images, which contains 5,896 instances of regular vehicles, 2,868 instances of pedestrians, 974 instances of vehicles in accidents, 706 instances of fires, and 418 instances of smoke.

When tested on 193 validation images, the model was able to achieve a mean average precision (mAP) at 50% Intersection of Union of 0.544 on accident vehicles, 0.717 on regular vehicles, 0.338 on pedestrians, 0.255 on fires, and 0.469 on smoke, for an overall model mAP of 0.465. Determining where fires ended and smoke began was very ambiguous, leading to low accuracy in those categories. Since the images gathered focused mainly on vehicles, pedestrians were often either in numerous crowds around an accident scene, or distant in the scene, resulting in low performance for pedestrians. To improve these results, a more comprehensive and balanced dataset of accident conditions and pedestrians must be gathered. Testing of the generated model on Raspberry Pi will be conducted to evaluate the runtime and memory usage.

†This work is supported in part by UNLV REU Site: Smart Cities (NSF grant no. 1950872).