Grant Bowlds

Stratification of Deep Neural Networks for CAVs

Mentors: Dr. Xenofon Koutsoukos, Dr. Martin Franzle, Dr. Willem Hagemann

1. General Problem and Context

Autonomous driving relies on computer vision to make most of its decisions. Whether a vehicle is accelerating, changing lanes, or avoiding pedestrians, connected and autonomous vehicles (CAVs) must first perform what is called object detection. This can be done in many ways, and most vehicles use light and radar (LiDAR) technologies alongside image processing to accomplish it. My research focused on image processing, which involves passing an image collected by the vehicle into a convolutional neural network (CNN) as a matrix of pixels and returning the category and location of every object within that image. Once the labels are returned, the software operating the vehicle can take appropriate action if it is needed. However, if the neural network cannot classify all the objects in the image correctly, there can be catastrophic errors. In a recent study by the University of Washington, industry standard CNNs were unable to detect stop signs that had been graffitied in one hundred percent of trials (Evtimov et al. 2017). While CAVs are still making this kind of mistake, they simply will not be safe enough to be on the road without human supervision. Consequently, improving the accuracy and speed of the CNNs that perform the object detection is one of the most important tasks to increase autonomy on the roads.

2. Description of the Specific Human-Cyber Physical System Problem

One major issue is that state-of-the-art object detection algorithms, such as Faster R-CNN, fail to recognize objects when snow is present in an image, even though the objects are still clearly visible to a human eye (Michaelis et al. 2019). This problem obviously can lead to significant complications with the software controlling the vehicle, and it can happen with any weather condition, not just snow. One simple solution to this problem is to create different CNNs for every foreseeable weather condition. Specialized networks almost always outperform general networks when they are trained on the same quality and amount of data. The issue with applying this concept to CAVs is that roads will rarely ever have the same driving conditions, and weather conditions are notoriously difficult to classify in an image. For example, although a road might be covered with snow, the weather could still be sunny, and a CNN could misclassify an image even though the road objects are more similar to those in snowy images. Because of this unreliability, even if a CAV has access to specialized networks, it may not be able to recognize which network is most optimal for that image. During this research project, I studied the effects of creating a new neural network that classifies weather conditions based on a triplet embedding process combined with a k-nearest-neighbors (k-NN) technique. The k-NN technique searches for the k closest embeddings in the dataset and generates a prediction based on the categories of these neighbors. The triplet network then has a threshold for minimum confidence and will return between one and the total number of possible weather classifications that meet that threshold set by the user. Once the triplet CNN returns this set, the image can be passed into all of the predicted stratified networks, and those networks can all predict what objects are present in that image.

3. Challenges of Reaching a Functional System

The biggest challenges of reaching a functional system are the amount of data available and the time needed to pass images through these networks. State-of-the-art neural networks have reached a point where they can perform almost as well as human eyes when they are given enough high-quality data to train on and enough time to process the image. I worked with the Berkeley Deep Drive 100k dataset for this project, the largest open-source image set for driving research, and still did not have enough images to achieve standard accuracy without overfitting to the training data. Another issue is computing time. Processing sizable images takes CPUs and even GPUs a considerable amount of time, and CAVs need to process images in real time. The time constraint means that images must be sized down to a fifteenth or more of their original size. For small objects and minute details (like snowflakes), this adjustment can cause errors and be difficult to circumnavigate.

4. Technical Problem and Research Setting

Due to the COVID-19 pandemic, the research took place remotely with regular phone calls and meetings with Dr. Xenofon Koutsoukos and Dimitrios Boursinos at Vanderbilt University and Drs. Martin Franzle and Willem Hagemann at the University of Oldenburg. The University of Oldenburg had created neural networks for object detection using the data from Berkeley Deep Drive 100k prior to the beginning of my project. I started my research by creating six different neural networks that were stratified based on a human-classified weather condition, imitating the work done at the University of Oldenberg. After these networks were built, I constructed an

oracle network to classify the weather condition in an image. After the base network was built, the weights from the training were loaded into the triplet network framework developed by Dimitrious Boursinos and Dr. Xenofon Koutsoukos (Boursinos and Koutsoukos 2020). This network consists of three copies of the same network that are joined together with an additional layer to generate a class prediction. The network takes three inputs: the base image, a positive anchor (another image of the same class), and a negative anchor (an image of another class). Initially, the network is trained by generating random triplet pairs. Once the network is trained, it can generate embeddings for final predictions. After a set of predictions is generated, the image will then be passed onto any of the six networks that the oracle believed it could belong to, and decision making will be based of the union of the objects detected, taking the safer path if there are any discrepancies.

5. Future Research

The goal of this research was to try to make existing CNN frameworks work as well as possible for object detection. The neural networks can be improved by continuously training them with new data as it becomes available. Additional research will also be done on the best way to upgrade the CNN framework itself. In a CNN, the base of a neural network is often represented as a graph or a tree, and training the networks sets weights on each layer of the underlying data structure. Every few years, new structures are studied and published, often with better performance than networks published in the past. In addition, the software should take the union of the object prediction sets and then find a way to safely classify disagreements in order for this solution to become fully functional.

References

Boursinos, Dimitrios and Xenofon Koutsoukos. (2020) "Trusted Confidence Bounds for Learning Enabled Cyber-Physical Systems." Workshop on Assured Autonomous Systems at the IEEE Symposium on Security and Privacy 2020. Available at: arXiv preprint arXiv:2003.05107.

Evtimov, ivan, Kevin Kykholt, Earlence Fernandes, Tadayoshi Kohno, Bo Li, Atul Parakash, Amir Rahmati, Dawn Song. (2017) "Robust physical-world attacks on machine learning models." Available at: https://s3.observador.pt/wp-content/uploads/2017/08/08133934/1707-08945.pdf.

Michaelis, Caludio, Benjamin Mitzkus, Robert Geirhos, Evgenia Rusak, Oliver Bringmann, Alexander S Ecker, Matthias Bethge, Wieland Brendel. (2019) "Benchmarking robustness in object detection: Autonomous driving when winter is coming." Available at: https://arxiv.org/abs/1907.07484.