HW: Build an AI Algorithm

An AI Method for Detecting Multimodal Street Use Robert Crump: 12231546

Executive Summary

The newly inaugurated Biden Administration has signaled support for large scale federal infrastructure spending, most notably, the incoming Secretary of Transportation, Pete Buttigieg. One very specific way to spend the $2 trillion planned for new infrastructure[1] that would combat climate change while revitalizing communities and spurring job growth is financing Vision Zero projects around the country. Producing good return on investment requires precision, and the deployment of an artificial intelligence (AI) could determine where money could be most effectively spent in support of multimodal street use. The AI method used for this project uses computer vision to detect pedestrians and cyclists actively using streets in an urban setting. The algorithm correctly identified the multimodal use % of the time, and unimodal use % of the time.

Background

Vision Zero is an infrastructure strategy originated in Sweden and promotes policies that aim to reduce traffic fatalities to zero[2]. Making streets safer produces many positive externalities related to social equity, sound municipal finance, and ecological preservation. Streets have a wide variety of uses among different citizens, and some need no changes to keep accommodating cars. Building new infrastructure where no pedestrians or cyclists travel would be an inefficient use of limited resources. Identifying which urban areas would most benefit from multimodal infrastructure could be predicted by traditional urban planning theories or could be informed by direct data gathering *in situ*. This project develops a proof of concept of how to design a computer vision tool that will identify where frictional interactions are most likely between vehicles and high vulnerability road users, i.e. cyclists and pedestrians.

The inputs of this project are images downloaded from Cityscapes Dataset. The output is a classification system that distinguishes urban landscapes from multimodal and unimodal uses. The intended user is a city planner, who gains the benefit of real world data to inform infrastructure investments.

Method

The dataset used to develop this tool are raw image files (portable network graphics format) downloaded from Cityscapes Dataset[3], a German research company. The images were

1. wsj.com/articles/buttigieg-pledges-to-support-bidens-2-trillion-infrastructure-plan-11611259580
2. visionzeronetwork.org/about/what-is-vision-zero/
3. https://www.cityscapes-dataset.com/file-handling/?packageID=3

sourced from a dashcam mounted on a vehicle travelling through Berlin in a single afternoon during light to medium traffic. Each image is the 20th in 30 frame video snippets recorded during the trip. There is no specific information about the speed of the vehicle, but I surmised from the environment and condition of the images that it reaches a top speed between 25-30 mph and a bottom speed of 0 when it comes to a complete stop.

To sort the 535 images in the Berlin repository, I copied all the images into folders labeled 'type1' for multimodal street use and 'type2' for unimodal street use. Then, I opened each folder and deleted images that I decided did not fit into the specification. I ended up with 273 multimodal images and 220 unimodal images for a total of 493. I reserved 27 and 22 images from the 'type1' and 'type2' folders respectively for testing and used the rest to train Google's Teachable Machine algorithm.

Unimodal use indicates that only vehicles are present in the roadway. Multimodal use is harder to define for computer vision because pedestrians and bicycles can be present in an image, but not actively using the roadway. To qualify multimodal use, I excluded motorbikes, bicycles separated from the roadway by a curb, and any pedestrian not within or on the edge of the roadway. I decided that it was important to include pedestrians waiting at crosswalks since they are about to cross the street even if they are not in it when the photo was captured. It's easy to distinguish this behavior as a human, but perhaps less straightforward for an AI.

Teachable Machine unique URL

Google Drive testing data folder

Results

The algorithm correctly identified multimodal use in 20 out of 27 instances and unimodal use in 17 out of 22 instances for a total of 36 / 49, or 73% accuracy. One potentially confounding factor is that the Teachable Machine crops images into a square, so the original rectangle format is not retained by the system. Pedestrians and cyclists on the periphery of some images have been cropped out. The AI is less sure about classifying images as 'type1' than 'type2'. Six instances in the multimodal set produced a less than 90% confidence in testing, while four instances in the unimodal set were below 90%. The AI struggled to identify bicycles that were

not in profile showing both wheels and pedestrians waiting to cross the street. Similarly, it failed to distinguish cyclists and pedestrians separated from the roadway by a curb.

Sources:

1. Mann, Ted. 2021. "Buttigieg Pledges To Support Biden'S $2 Trillion Infrastructure Plan ". WSJ. https://www.wsj.com/amp/articles/buttigieg-pledges-to-support-bidens-2-trillion-infrastructure-plan-11611259580.
2. "What Is Vision Zero?". 2021. Visionzeronetwork.Org. https://visionzeronetwork.org/about/what-is-vision-zero/.
3. M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.