Detection of Active Emergency Vehicles using Per-Frame CNNs and Output Smoothing

Meng Fan, Craig Bidstrup, Zhaoen Su, Jason Owens, Gary Yang, Nemanja Djuric
Aurora Innovation, Inc.

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Motivation

While inferring common actor states (such as position or velocity) is an important and well-explored task of the perception system aboard a self-driving vehicle (SDV), it may not always provide sufficient information to the SDV. This is especially true in the case of active emergency vehicles (EVs), where light-based signals also need to be captured to provide a full context. We consider this problem and propose a sequential methodology for the detection of active EVs, using an off-the-shelf CNN model operating at a frame level and a downstream smoother that accounts for the temporal aspect of flashing EV lights. We also explore model improvements through data augmentation and training with additional hard samples.

Data set and labeling

Category	Distribution
Time of logs	Day: 81.1%, Night: 18.9%
Vehicle type	EV: 3.4%, non-EV: 96.6%
EV type	police vehicle: 80.0%, fire vehicle: 13.4%, ambulance: 6.6%
EV activeness	active: 90.0%, inactive: 10.0%
Bulb state of active EVs	bulb-on: 91.8%, bulb-off: 8.2%

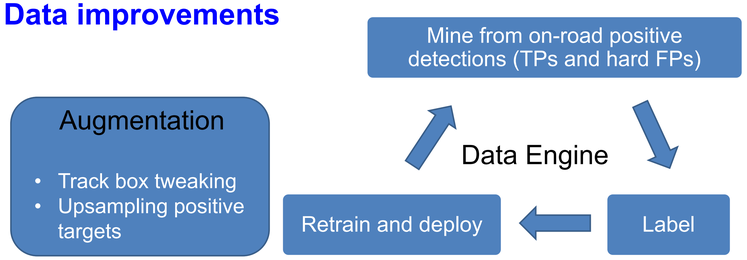
- On-road data collection using autonomous driving fleet
- ~12000 vehicles in total
- 4D+2D labeling and association
- EV attribute labeling
  - Is it an EV?
    - If so, what category is it?
    - If so, is the EV active?
      - If so, is it bulb on at the current frame?

Methodology

- The EV detection system
  - Can handle hundreds of actors (processed as a batch) in real-time, with an average latency of less than 10 milliseconds
  - Preprocessing: track projection onto image; crop and resize
  - Inference
    - Input: batched image crops
    - EV Classifier: ResNet-18 backbone + a single head
    - Output: probabilities indicating whether a certain vehicle is an active EV or not
  - Postprocessing
    - The smoother is a cyclic buffer that keeps a ledger of the last 25 valid EV outputs for each vehicle track's history and outputs a smoothed result.
    - We require that there are at least 6 detected frames of the actor and more than T = 50% of per-frame active EV outputs in the buffer to mark the vehicle as being an active EV. This helps suppress transient positive outputs and thus mitigate false positives (FPs)

Example detections (top row: true positives, bottom row: difficult true negatives); All correctly classified by our model.

Data improvements



Results

Per-frame classifier results (A: data augmentation; M: mined data from data engine)

Model	% change of max-F1	% change of precision at 0.8 recall
Baseline	0	0
+ A	3.11	5.3
+ A + M	5.73	10.4

Per-actor results of the output smoother

Smoother threshold T	% change of precision	% change of recall	% change of F1 score
0%	0	0	0
30%	10.30	-0.68	5.07
50%	9.87	-0.51	4.87
70%	6.80	-20.0	-7.02

Summary

We considered the problem of detecting active emergency vehicles in the context of self-driving vehicles. To address this task we proposed to use a frame-level EV detector whose outputs are fed to an output smoother, which captures the temporal dimension of such actors. Finally, we evaluated the method on large-scale, real-world data, and made further improvements by data augmentation and by training with additional hard samples mined using a data engine.