Tracking Vehicles Across Multiple Viewpoints

Paul Warren, pwarren@stanford.edu

Introduction

Visual query system for vehicle re-identification

- Input: image of vehicle
- Output: N images of the same vehicle

VeRi dataset examples

Dataset

VeRi [1]. 50,000 images from ~20 nearby security cameras in China. Each car imaged multiple times by multiple cameras.

Common Approaches

- Vehicle embeddings (like word2vec)
- Vehicle attributes (like license plate, make, model)
- Detection attributes (like where and when we detected the vehicle)

My Experiment

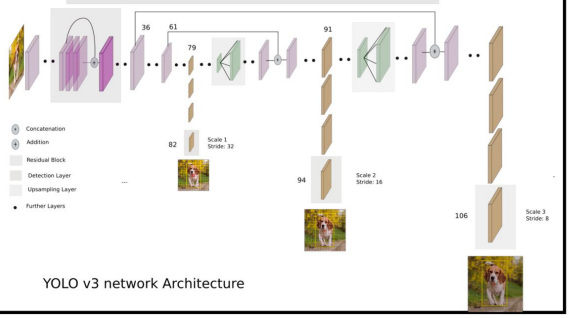
Most approaches have 2 models:

- Object detection
- Embedder

Can we use the object detection bounding box feature vector as the embedding?

Model: YOLOv3 [2] trained on MS-COCO. (80 classes, ~385k images), length 255 feature vector

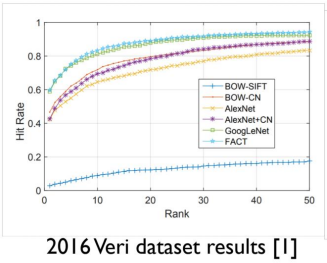
YOLOv3



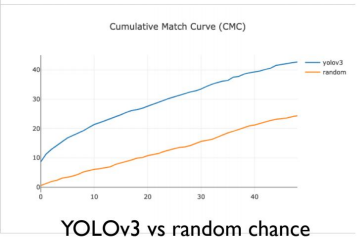
YOLO v3 network Architecture

Results

Twice as good as random chance, half as good as 2016 state-of-the-art "Cumulative matching characteristic".



2016 Veri dataset results [1]



YOLOv3 vs random chance

Conclusion and Future Work

- Just YOLOv3 actually works some of the time
- Worth investigating further!

Future Work

- Further train YOLOv3 on surveillance data
- Try embeddings from Faster-RCNN, Mask-RCNN
- More visual inspection of results

References

[1] X. Liu, W. Liu, H. Ma and H. Fu, "Large-scale vehicle re-identification in urban surveillance videos," 2016 IEEE International Conference on Multimedia and Expo (ICME), Seattle, WA, 2016, pp. 1-6.

[2] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.

…many more in paper