Comparing Object Detection Algorithms

Ryan Vavruska

The Problem

- With the fast advancing expertise of Object Detection, the best algorithm to use for certain situations is constantly changing
- Need to adapt and change based on what is currently best for the task
- Systems that run in real-time need to be quick
- Looking to compare three different learning algorithms to see which one performs the best on a data set.
  - YoloV5
  - SSD
  - Faster RCNN

Approach

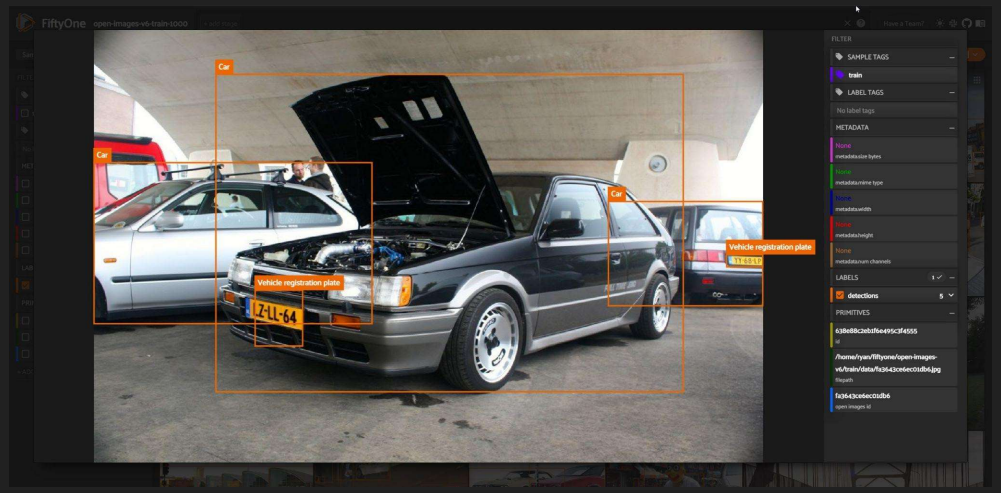
- Compare SSD, R-CNN and YOLOv5 using OpenImagesV6.
- Single Shot Multibox Detector (SSD) - Object detection algorithm that relies on the generation of bounding boxes and the extraction of feature maps.
  - Built on a feed forward complex network that builds an amount of standard-size bounding boxes and for each object that in contained within these boxes they are given a score
- Region-based Convolutional Neural Network (R-CNN) - Object detection algorithm that combines region proposals for object segmentation and high capacity CNNs for object detection.
  - Uses selective search algorithm which proposes candidate regions.
  - Proposals are wrapped and extracted for distinct features which are then fed into an SVM for recognizing objects of interest in the proposal regions.
- Faster R-CNN - A R-CNN where the proposals are learned by the network, eliminating the need for a selective search algorithm
- YOLOv5 (You Only Look Once) - Considered one of the best object detection algorithms available. Utilizes a highly modified architecture named Darknet.

Approach

- Each algorithm was implemented in Python using PyTorch as a framework
- A subset of the dataset was taken that included a single class, Cars.
- Average precision was used as the primary means of comparing networks, with the mAP equalling the AP since there is only one class.
- Models were trained using Google Cloud Platform on a T4 GPU.

Dataset

- Google Open-Images-V6
  - Dataset containing approximately 9 million annotated images.
  - These images contain complex scenes often containing 8 or more images.
  - Annotated by humans to ensure accuracy and consistency in the annotations.
- Images resizes to 416x416 to standardize them



Yolov5

- Model pre-trained on the Microsoft COCO dataset
- Three different models trained
  - All layers trained
  - Last 14 layers trained, first 10, the backbone, frozen
  - All layers but output layer frozen
- Model with only the backbone frozen performed the best
- Produced an AP value of 0.6993

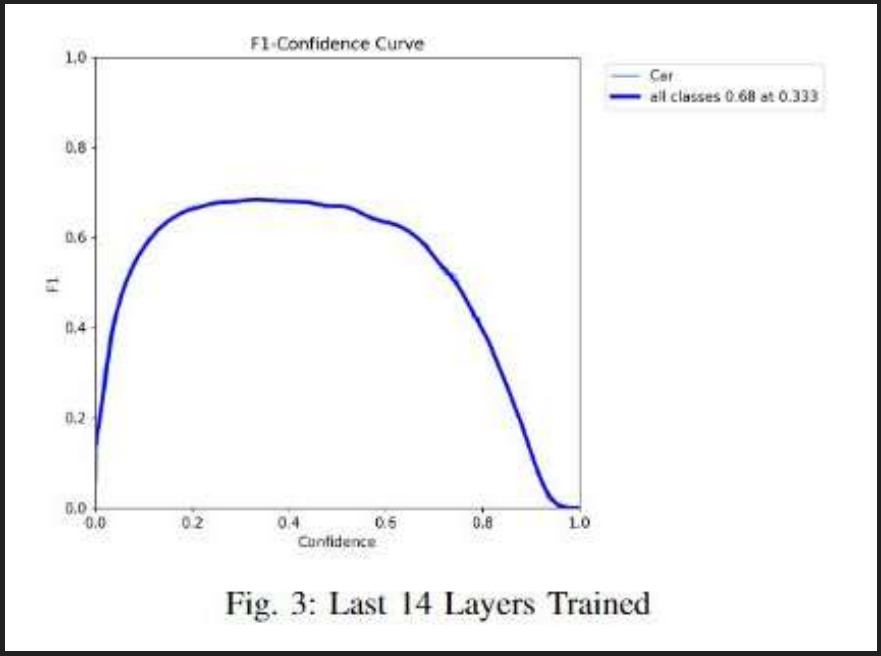


Fig. 3: Last 14 Layers Trained

Yolov5

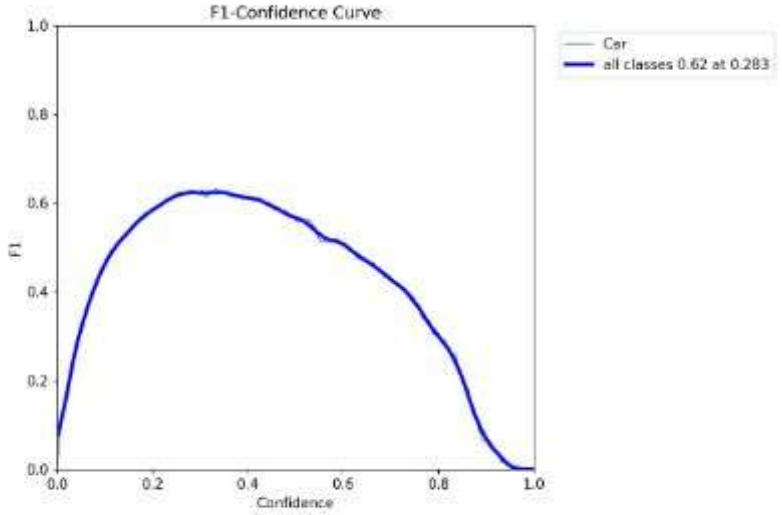


Fig. 2: All Layers Trained

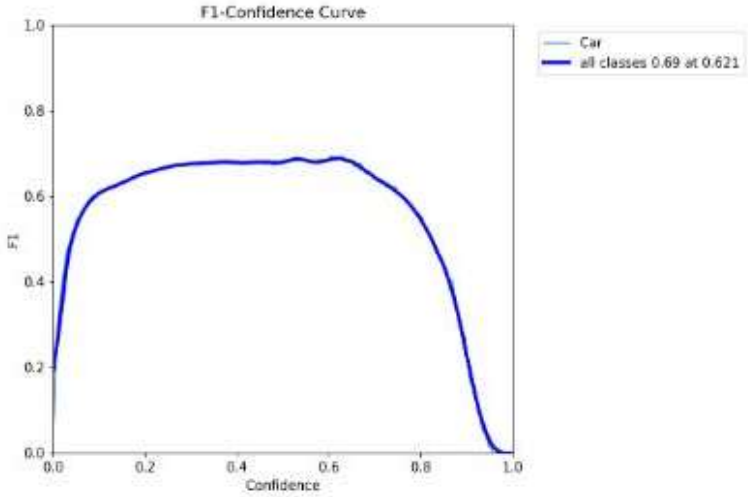


Fig. 4: Output Layer Trained

Yolov5 Output Map

Single Shot Multibox Detector (SSD)

- Model pre-trained using Microsoft COCO
  - Model trained on Car subset in the VOC format.
- Produced an AP of 0.498

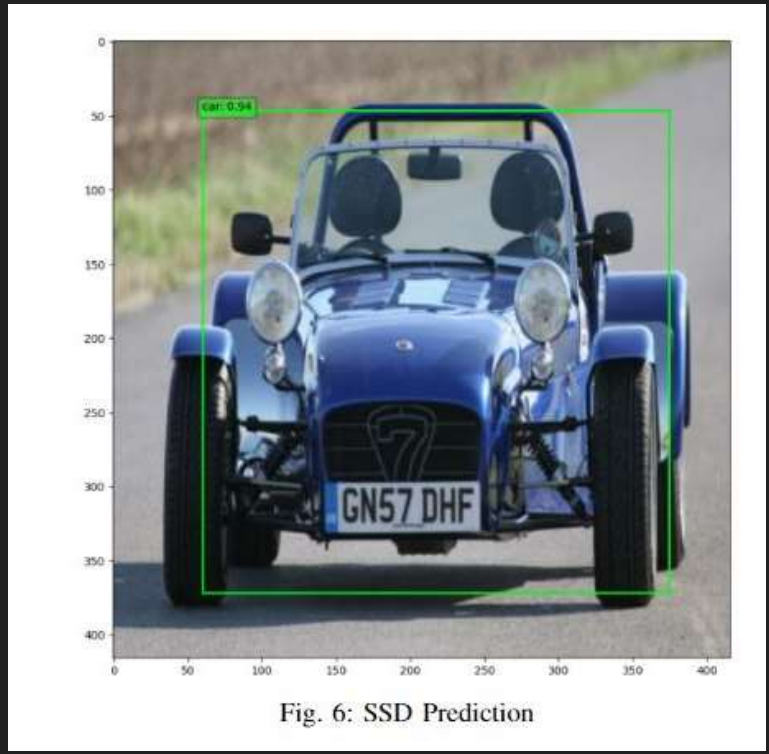


Fig. 6: SSD Prediction

Faster RCNN

- Used a pre-trained version of PyTorch's fasterrcnn_resnet50_fpn.
- Model did not properly acclimate to the dataset
- No AP worth capturing as can be seen in the image to the right.

Fig. 7: Faster RCNN Prediction

Analysis

- Yolov5 performed the best among the three different algorithms.
  - Shows great performance and would likely be the best in terms of real-time systems.
- SSD falls slightly behind Yolov5 when it comes to accuracy

Algorithm Name	mAP0.5
Yolov5	0.6993
SSD	0.498
Faster RCNN	-

References

[1] J.C. Nascimento and J.S. Marques. “Performance eval-uation of object detection algorithms for video surveil-lance”. In: IEEE Transactions on Multimedia 8.4 (2006),pp. 761–774. DOI: 10.1109/TMM.2006.876287.

[2] Wei Liu et al. “SSD: Single Shot MultiBox Detector”. In: Computer Vision – ECCV 2016. Ed. by Bastian Leibeet al. Cham: Springer International Publishing, 2016,

pp. 21–37. ISBN: 978-3-319-46448-0.

[3] Mohammed Noman, Vladimir Stankovic, and Ayman Tawfik. “Object Detection Techniques: Overview andPerformance Comparison”. In: 2019 IEEE Interna-tional Symposium on Signal Processing and Information Technology (ISSPIT). 2019, pp. 1–5. DOI: 10 . 1109 / ISSPIT47144.2019.9001879.

References

[4] Glenn Jocher. YOLOv5 by Ultralytics. Version 7.0. May 2020. DOI: 10.5281/zenodo.3908559. URL: https://github.com/ultralytics/yolov5.

[5] Alina Kuznetsova et al. “The Open Images Dataset V4:

Unified image classification, object detection, and visual

relationship detection at scale”. In: IJCV (2020).

[6] Shrey Srivast et al. Comparative Analysis of Deep Learn-ing Image Detection Algorithms. Dec. 2020. DOI: 10 .21203/rs.3.rs-132774/v1.

References

[7] Raza Rizwan. google-open-image-cars-dataset Dataset. https://universe.roboflow.com/raza-rizwan/google-open- image - cars - dataset. Open Source Dataset. visited on 2022-12-05. July 2022. URL: https://universe.roboflow.com/raza-rizwan/google-open-image-cars-dataset.

[8] Max deGroot. SSD: Single Shot MultiBox Object Detector, in PyTorch. URL: https://github.com/amdegroot/ssd.pytorch.