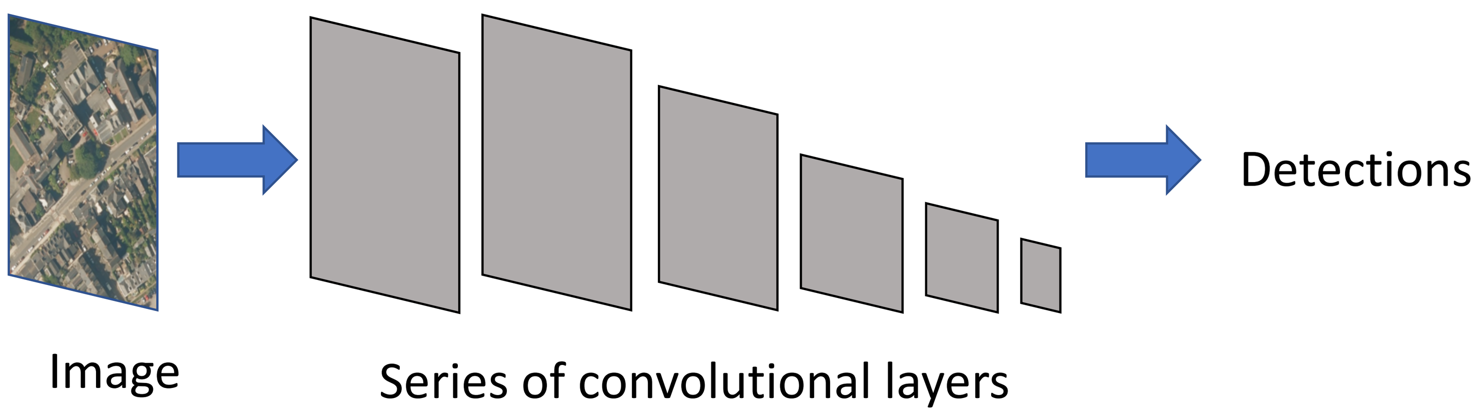
Vehicle detection in aerial imagery using TensorFlow

Chloe Steele
Geographic Data Science Lab, Department of Geography and Planning, University of Liverpool
C.M.Steele@liverpool.ac.uk

Introduction

Vehicle detection from aerial or satellite imagery is an important area of research and is used in applications such as vehicle tracking, traffic surveillance, intelligent transport systems and traffic flows (Yang et. al., 2018).

This research involves the use of convolutional neural network (CNN) object detection models on vehicles contained in aerial images. Specifically, the Single Shot Detector (SSD) model and Faster Region Convolutional Neural Network model (Faster R-CNN) are looked at to determine how accurate they are at detecting vehicles.



Simple diagram of CNN model.

Aims

- Train two different convolutional neural network object detection models on UK aerial imagery.
- Compare the performance of the two models at detecting vehicles in unseen aerial images.
- Determine whether these models can be used for monitoring traffic flows and looking at car park capacities.

Data

Models

Pre-trained models were obtained from the TensorFlow Object Detection API GitHub repository which is an open-source framework (Huang et. al., 2017). Both models are pre-trained on the COCO image dataset.

Aerial images

The images used to train the models were created and licensed by Getmapping plc., and downloaded from EDINA Digimap.

- 25cm resolution vertical ortho-photography.
- 18 aerial images used for model training.

Methods

Image labelling

- LabelImg software (Tzutalin, 2015) used to manually draw ground truth boxes around all vehicles in each image and assign the label of 'car'.
- 14 images containing 519 tagged vehicles used as 'train' set, while 4 images containing 167 tagged vehicles as the 'test' set for training.

Model training

- Faster R-CNN trained on the manually labelled aerial images for 15 hours, reaching Total Loss of 0.09 after 9276 steps.
- SSD trained on labelled aerial images for 7.5 hours, reaching Total Loss of 0.25 after 342 steps.

Results

- All predictions over 80% accurate on unseen images.
- SSD identifies 'false positives' unlike Faster-RCNN.
- Both models less accurate at detecting vehicles in car parks.

Conclusions

- Faster R-CNN has a higher success rate and does not falsely identify objects as vehicles, despite longer training time.
- Vehicle detection in car parks is less successful than on roads.
- Models should be trained longer to increase overall accuracy.

References

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S. and Murphy, K. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [online] IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Available at: https://arxiv.org/abs/1611.10012 [Accessed 9 May 2019].

Tzutalin, LabelImg, (2015). GitHub Repository available at: https://github.com/tzutalin/labelImg

Yang, C., Li, W. and Lin, Z. (2018). Vehicle Object Detection in Remote Sensing Imagery Based on Multi-Perspective Convolutional Neural Network. *ISPRS International Journal of Geo- Information*, 7(7), p.249.