Heidi M Hurst heidimhurst.github.io

Supervisor: Dr. Santosh Bhattarai

MSc Geographic Information Science 2018
UCL Department of Civil, Environmental and Geomatic Engineering, Gower St, London, WC1E 6BT

Impact of resolution on car detection accuracy in satellite imagery using neural networks

UCL

Research Question

How do spatial resolution and object resolution (pixels per car) impact small car detection accuracy in satellite imagery using convolutional neural networks (CNNs)?

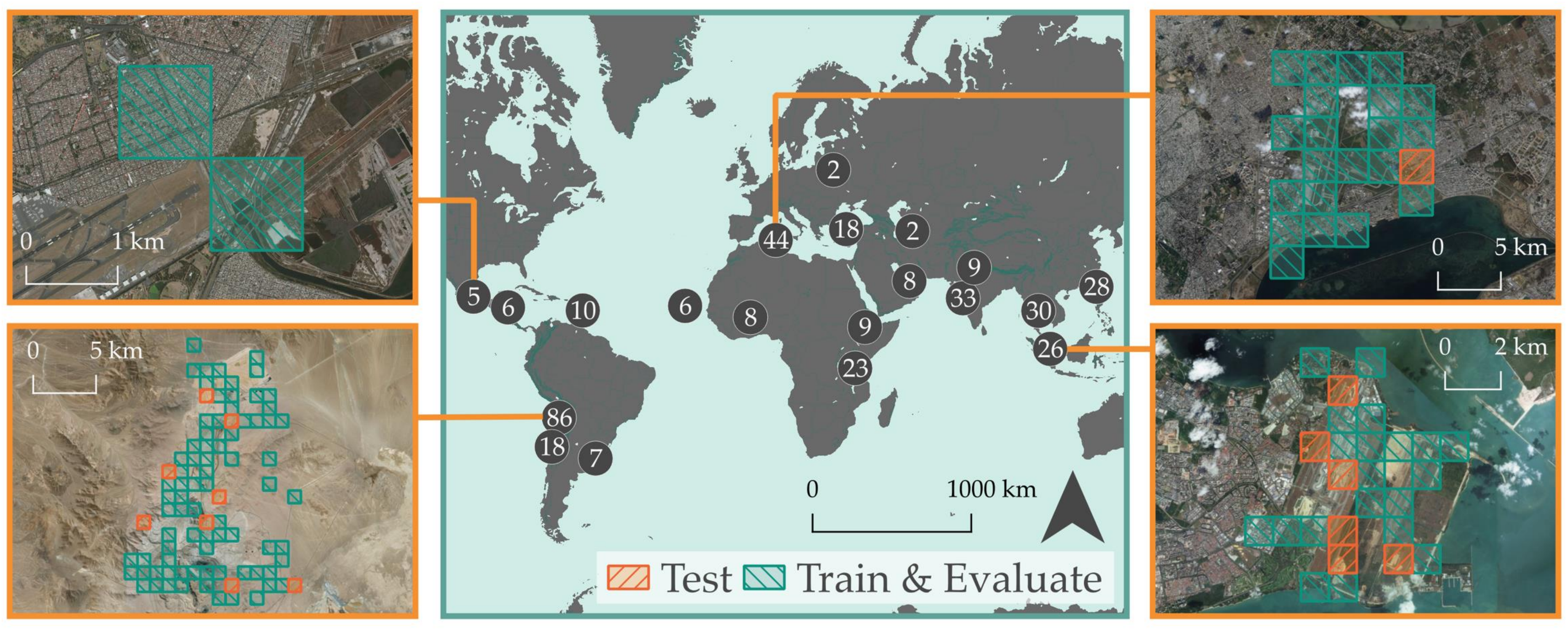

Figure 1: Four sites showing breadth of imagery context. Numbers on the world map show xView image quantity by location (ESRI World Imagery).

Methods

For this study, 14 different Inception Single Shot Multibox Detector (I-SSD) CNNs were trained on the xView overhead imagery data set [1].

Two separate trials were conducted. Each trial consisted of training and assessing the model at each of seven downsample factors (df), where df 2.0 reduces image size by half in both height and width.

1. **Prepare data**: xView data is either (a) downsampled by the given df (reducing image size to mimic low spatial resolution imagery) or (b) upsampled after being downsampled (returning image to the original object resolution while decreasing spatial resolution to simulate low res imagery that has been made larger; see Figure 2).
2. **Train & evaluate model**: Each model is trained for 100,000 steps and continuously evaluated on a subset of the data using the TensorFlow Object Detection API on a GPU enabled remote device.
3. **Export model**: Training is complete at the step where evaluation loss reaches a minimum. Models are then exported for future use.
4. **Run inference**: Saved models are used to predict car bounding boxes on images the model has never seen (see Figure 3) .
5. **Assess model accuracy**: Average precision (AP), a metric of model accuracy and quality, is computed for each trial (see Figure 4).

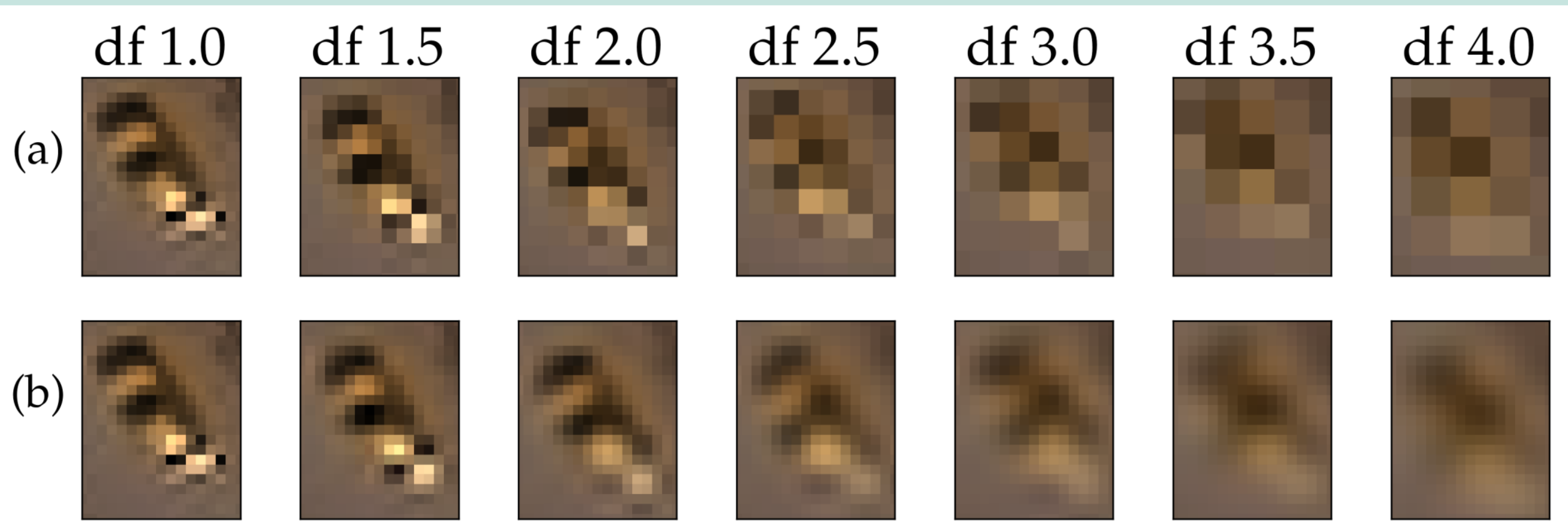

Figure 2: Example of a yellow car (a) downsampled and (b) subsequently upsampled at each of seven dfs creating a blur effect due to information loss.

Motivation

Identification of small cars has a wide variety of civilian and military applications, including understanding population flow after a disaster. Without automation, cataloguing these objects globally is impossible.

Results

Models trained on upsampled imagery had significantly higher AP than those trained on downsampled images at the same df (see Figure 4). AP is more strongly correlated with median object size than with df (standard correlation coefficient 0.90 vs -0.65). Overall, decreasing either resolution leads to decreased AP (see Figure 5).

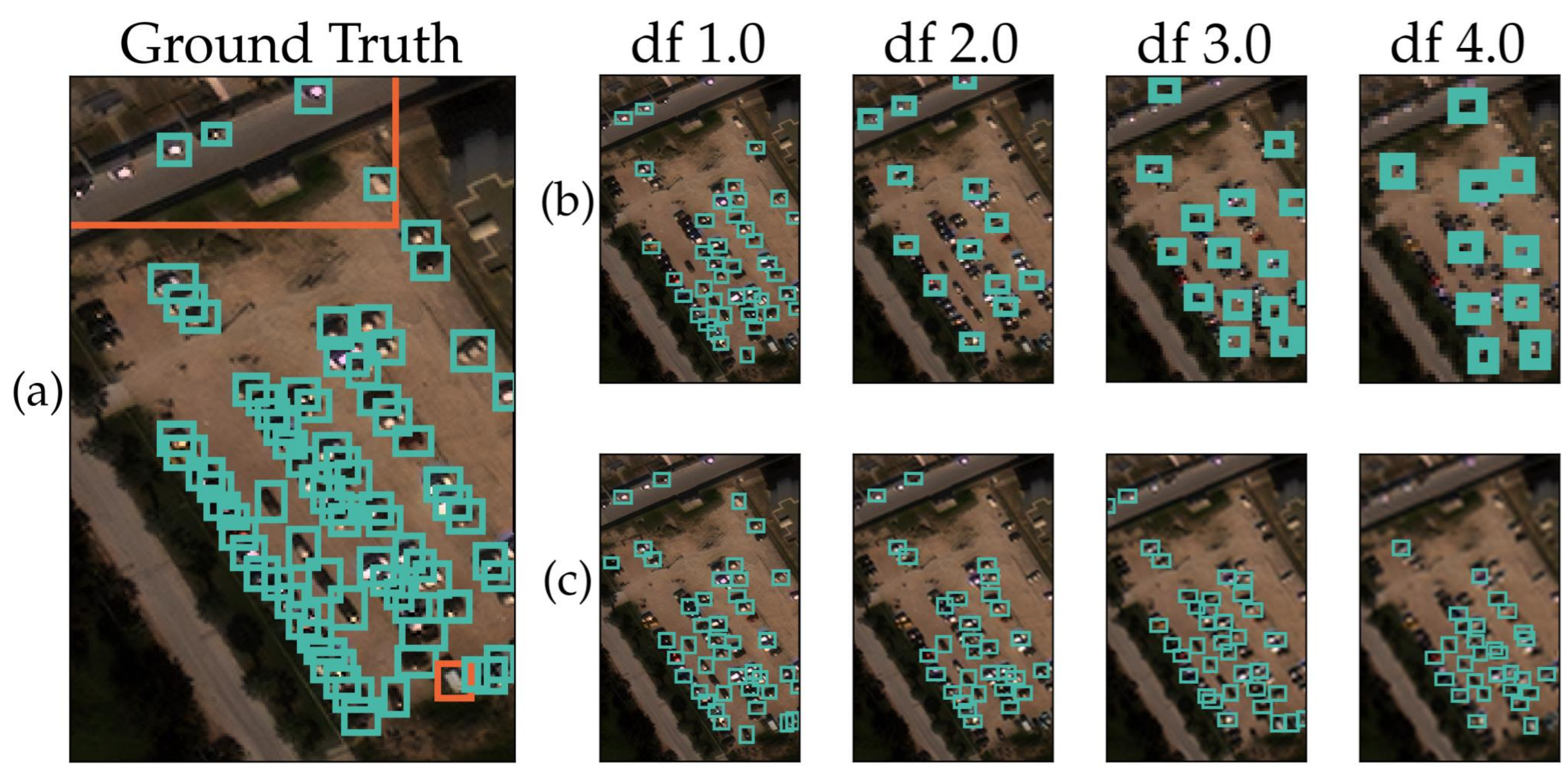

Figure 3: (a) ground truth boxes with non car classes shown in orange. Example performance of models trained on (b) downsampled and (c) upsampled images.

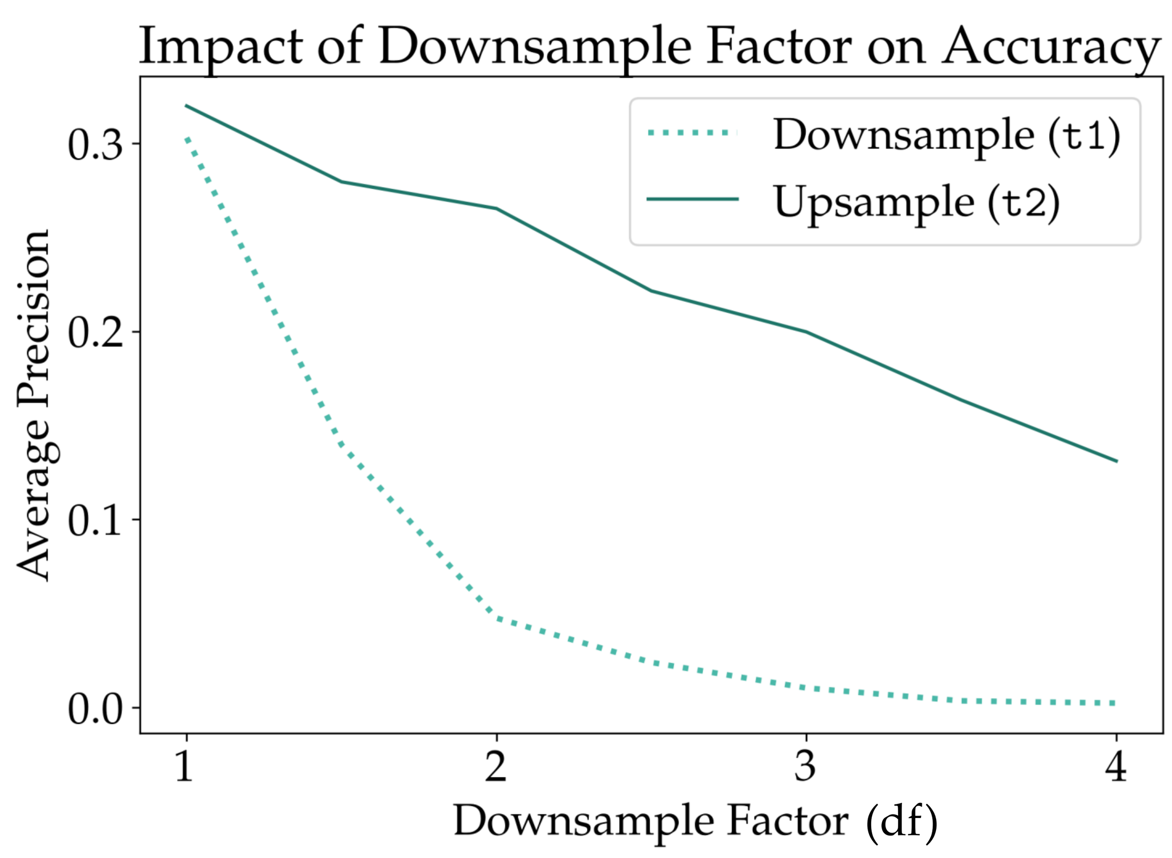

Figure 4: Raw results for each trial showing impact of resolution on accuracy (higher df means lower spatial resolution).

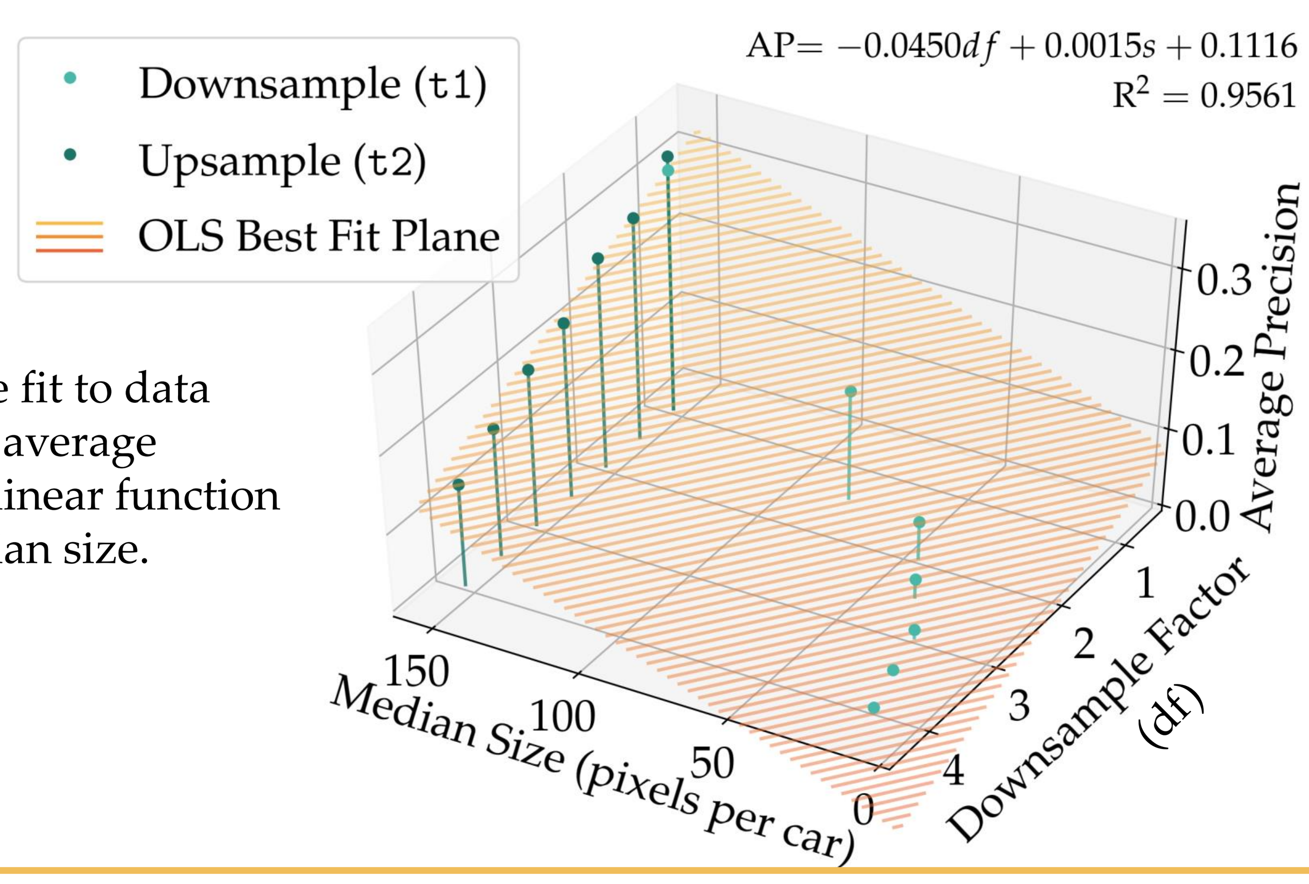

Figure 5: Plane fit to data points models average precision as a linear function of df and median size.

Reference: [1] Lam, Darius et al. (2018). "xView: Objects in Context in Overhead Imagery". In: *arXiv:1802.07856 [cs]*. arXiv: 1802.07856. URL: http://arxiv.org/abs/1802.07856.

Conclusions

- Models trained on low resolution imagery may be improved by upsampling the imagery to increase object resolution.
- Object resolution may contribute more to average precision than spatial resolution.
- Decreasing either spatial resolution or object resolution leads to decreased accuracy.
- Average precision was modeled as a linear function (plane) of df and median size (see Figure 5).
- Future research should investigate how this relationship is affected by quantity of example image or object type.