Satellite Image Object Detection with Mask R-CNN

Chase Toyofuku-Souza
2296478

Identifying the Problem

- Large amount of satellite imagery data available
- Takes a long time to click through
- Looking to buy a new house
- It’s not fun to use existing applications such as Zillow, so I built my own CNN to identify possibly affluent/cheaper areas to live in

Importance/Other applications

- Educated on the area surrounding future home
- More swimming pools and cars typically means more expensive areas
- Confirm house prices
- Flying drones over city
    - Detect stolen cars
    - Houses/forests on fire

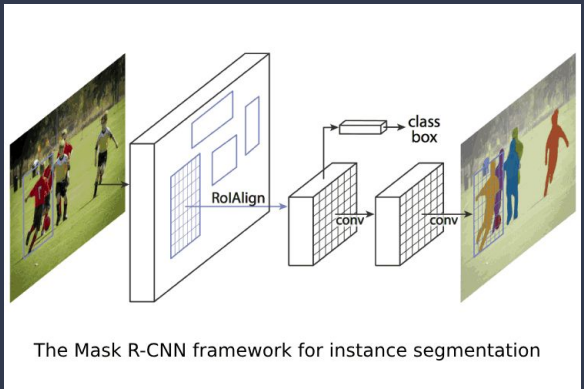
Data

- https://www.kaggle.com/kbhartiya83/swimming-pool-and-car-detection
- 2703 test images
- 3748 train images/labels
- Two classes (swimming pool, cars)
- All in same geographic area, around same time of day
- Average of 4.31 objects/image
- 13022 cars, 3150 swimming pools

Sample Data

```
<size>
    <width>224</width>
    <height>224</height>
    <depth>3</depth>
</size>
<object>
    <name>1</name>
    <bndbox>
        <xmin>58.47</xmin>
        <ymin>152.31</ymin>
        <xmax>69.58</xmax>
        <ymax>163.43</ymax>
    </bndbox>
</object>
<object>
    <name>1</name>
    <bndbox>
        <xmin>10.32</xmin>
        <ymin>205.68</ymin>
        <xmax>21.43</xmax>
        <ymax>216.80</ymax>
```

Model



The Mask R-CNN framework for instance segmentation

- Mask R-CNN
- Instance segmentation
- Separate different objects in a picture/video
- Generates proposals where regions may be
- Predicts class, refines bounding box, generates a mask

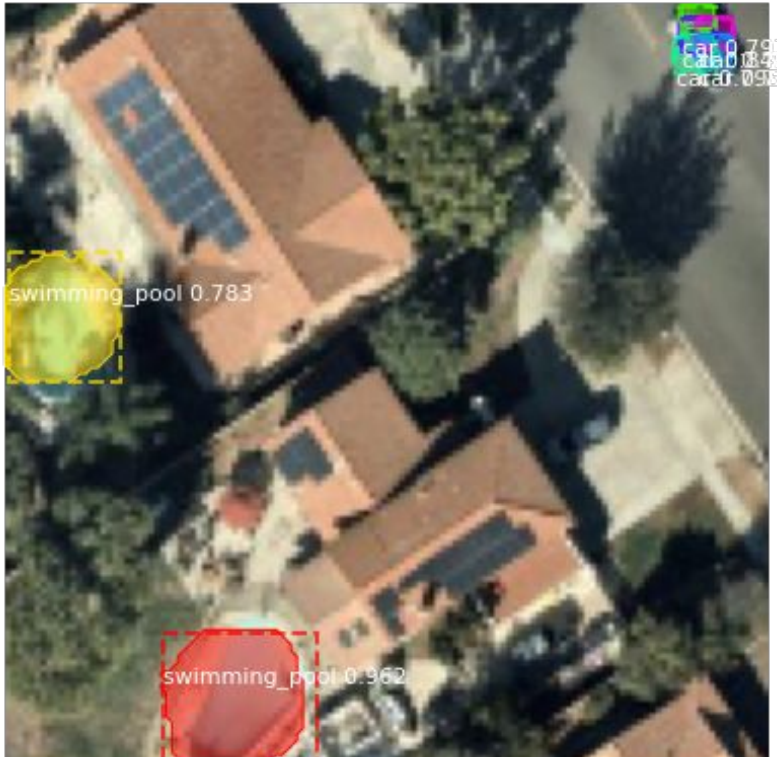
Model

- LR of .001, fine-tuned with LR of .0001
- Image shape of 128 x 128 x 3
- 75 Train ROI’s per image
- COCO used for initial weights
- No dropout/regularization

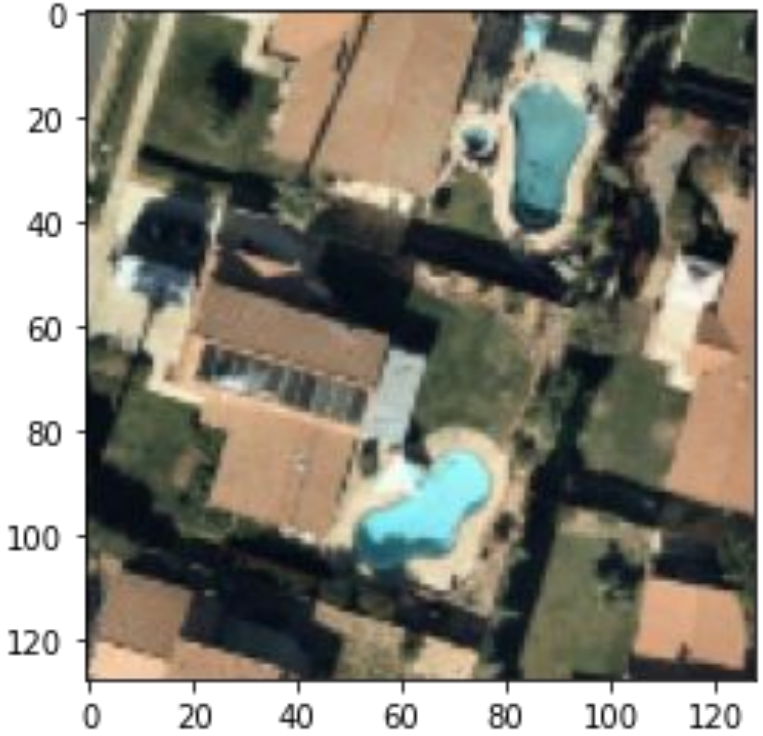
BBox Demonstration

BBox Demonstration

Model Results

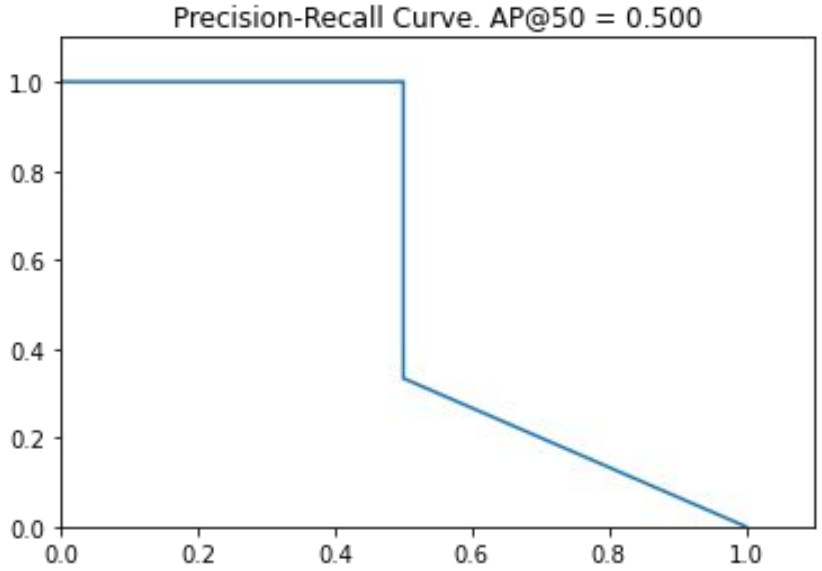


Model Results



Model Results

- Loss: 1.61
- mAP: 0.255
- Average F1: 0.211
- Average recall: 0.33



Model Results

- Shows that it’s possible to detect multiple classes/objects using low quality images
- Can be applied to classifying other objects
- Looking at too many regions
- Low F1 score indicates that model is probably overfit
- Feed it more training data
- Lower number of epochs
- Use different initial training weights

Citations

- https://prnewswire2-a.akamaihd.net/p/1893751/sp/189375100/thumbnail/entry_id/1_su9da4fu/def_height/1001/def_width/1911/version/100011/type/2/q/100
- https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd
- https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e
- https://github.com/matterport/Mask_RCNN
- https://www.edureka.co/blog/tensorflow-object-detection-tutorial/
- https://machinelearningmastery.com/how-to-train-an-object-detection-model-with-keras/
- https://blog.paperspace.com/mask-r-cnn-tensorflow-2-0-keras/
- https://github.com/matterport/Mask_RCNN/issues/2165
- https://stackoverflow.com/questions/59308263/using-a-tf-tensor-as-a-python-bool-is-not-allowed-in-graph-execution-use-ea
- https://stackoverflow.com/questions/55149026/tensorflow-2-0-do-you-need-a-tf-function-decorator-on-top-of-each-function