# Detecting objects with high object region percentage

Fen Fang , Qianli Xu, Liyuan Li, Ying Gu and Joo-Hwee Lim

*Department of Visual Intelligence , Institute for Infocomm Research, Singapore*

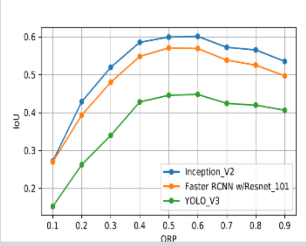
## Introduction

Object shape is a subtle but important factor for object detection. It has been observed that the object-region percentage (ORP) can be utilized to improve detection accuracy for elongated objects, which have much lower ORPs than other types of objects. In this paper, we propose an approach to improve the detection performance for objects with high ORPs.

- We provide a straightforward method to improve object detection performance on high-ORP objects. Our method is able to achieve the state-of-the-art performance without using complicated network structures.
- We develop an algorithm to uniformly adjust bounding boxes of objects to achieve the optimal ORP, and a DCNN to learn enlarging ratios toward different directions to fit the detected bounding box with the object.
- Our method provides insight into the detection of small objects, which is a challenging task for general object detectors. We show that expanding the bounding box helps to mitigate feature loss due to convolution operations, which improves the recall of small objects.

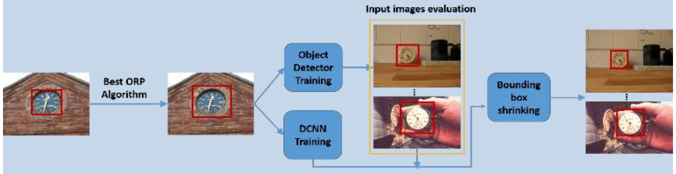
## Motivation

Objects with high ORPs, while the SNR is high, the detection performance is nevertheless lower than objects with intermediate ORPs. Empirical results from multiple detection frameworks have shown that there is an optimal ORP range (0.4 0.6), where detection performance is higher than the two extremities.




(a) (b) (c)

## Methodology




- ➢ Enlarge the bounding box around its centre by a certain ratio to left, right, top and bottom uniformly to achieve the optimal ORP.
- ➢ We employ a Faster R-CNN with *Resnet*-101 as the backbone, To train an object detector on the optimized bounding boxes.
- ➢ DCNN to learn the ratios for shrinkage toward four directions to shrink the predicted bounding box back to fit the object as tight as possible.
- ➢ With predicted enlarging ratios toward 4 directions from trained DCNN model, we can shrink predicted bounding box of object to get more accurate location.

## Results

### Quantitative Results on COCO Validation Set

| Class Name | Baseline (Faster rcnn-res101) | | | EfficientDet_d7 | | | Our Approach_A | | | Our Approach_B | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AP@0.5 | AP@0.75 | mAP | AP@0.5 | AP@0.75 | mAP | AP@0.5 | AP@0.75 | mAP | AP@0.5 | AP@0.75 | mAP |
| **apple** | 0.2302 | 0.2075 | 0.1852 | 0.3422 | 0.2809 | 0.2760 | **0.3734** | **0.3135** | **0.2932** | 0.3625 | 0.2877 | 0.2786 |
| **bowl** | 0.4636 | 0.4303 | 0.3821 | 0.5603 | 0.5430 | 0.5186 | 0.6317 | **0.5949** | **0.5332** | **0.6321** | 0.5625 | 0.5205 |
| **clock** | 0.7310 | 0.7218 | 0.6549 | 0.8181 | 0.7812 | 0.7440 | 0.8238 | **0.8107** | **0.7568** | **0.8371** | 0.7936 | 0.7467 |
| **donut** | 0.4278 | 0.4066 | 0.3784 | 0.6155 | 0.5892 | 0.5650 | **0.6636** | **0.6387** | **0.5940** | 0.6577 | 0.6391 | 0.5801 |
| **frisbee** | 0.7229 | 0.7229 | 0.6980 | 0.8758 | 0.8758 | 0.8618 | **0.9001** | **0.9001** | **0.8624** | 0.8745 | 0.8745 | 0.8509 |
| **orange** | 0.3545 | 0.3121 | 0.2977 | 0.3958 | 0.3352 | 0.3396 | 0.4910 | **0.4616** | **0.4242** | **0.4928** | 0.4062 | 0.3577 |
| **oven** | 0.5578 | 0.3837 | 0.3590 | 0.6550 | 0.5410 | 0.4994 | 0.6969 | **0.5567** | **0.5103** | **0.6977** | 0.5304 | 0.4987 |
| **paking_meter** | 0.5494 | 0.5495 | 0.4821 | 0.6651 | 0.6248 | 0.5976 | 0.8154 | **0.8154** | **0.7408** | **0.8179** | 0.7924 | 0.6576 |
| **sports_ball** | 0.4999 | 0.4999 | 0.4917 | **0.7188** | **0.7188** | **0.7101** | 0.5941 | 0.5909 | 0.5730 | 0.6042 | 0.5997 | 0.5583 |
| **stop_sign** | 0.6213 | 0.6213 | 0.5984 | 0.7775 | 0.7775 | **0.7723** | **0.8010** | **0.8010** | 0.7699 | 0.7726 | 0.7372 | 0.7139 |
| **toaster** | 0.4769 | 0.4769 | 0.4056 | 0.5476 | 0.5476 | 0.5049 | **0.7026** | **0.7026** | **0.6278** | 0.6994 | 0.6349 | 0.5917 |
| **TV** | 0.6899 | 0.6501 | 0.5703 | 0.7500 | 0.7438 | 0.6768 | 0.7908 | **0.7759** | **0.6841** | **0.8086** | 0.7558 | 0.6788 |

we select 12 classes from the COCO dataset based on the statistics of ORP distribution on COCO training set and common sense of object appearance. We compared our method with two methods namely, (1) an off-the-shelf Faster R-CNN with ResNet 101 (baseline) and (2) the state-of-the-art object detection method, EfficientDet d7 on COCO validation set containing 5000 images. It can be observed from the table that our Approach achieve much higher performance than Baseline, comparing with EfficientDet d7, our Approach perform better in 10 out of 12 classes.

### Qualitative Results on COCO Validation Set

***Ground Truth***


***Our Results***


## Conclusions

In this paper, we have proposed an approach to improve performance of object detection task on high-ORP objects. We first enlarge the bounding box of object to obtain the optimal ORP for Faster R-CNN detector training. Then, we train a DCNN for enlarging ratios estimation towards different directions. Finally, we shrink back detected bounding box of an object based on estimated enlarging ratios in the evaluation stage. In the future, we plan to design a shape-sensitive object detector that automatically adapts its behaviour for objects with varying ORPs.

## Acknowledgement

**This research is supported by the Agency for Science, Technology and Research (A*STAR) under its AME Programmatic Funding Scheme (Project A18A2b0046).**