UFO RPN: A Region Proposal Network for Ultra Fast Object Detection

Wenkai Li , Andy Song
RMIT University

Abstract

- Object detection is widely used in many industrial applications.
- Current object detection algorithms require high computational cost.
- We focus on the region proposal stage to reduce the number of candidate object regions sent to the detector.
- Our study shows that high resolution input is not a must for high accuracy. The use of down-sampled images can further reduce computation costs while retaining or even improving accuracy.
- The class IoU for MS COCO subsets achieves 40% to 70% and the inference speed on GTX 1080Ti can achieve above 1000 FPS performance.

Related Work & Motivation

- The generation of region proposals is an important step in object detection. It determines how the following expensive classification network is used.
- Inspired by the selective visual attention mechanism of human vision mechanism, saliency detection techniques can be used to generate region proposals with background areas somewhat removed [1].
- Salience information is preserved in low resolution gray scale (LG) images, according to human eye-tracking and computational modelling experiments [2].
- Different from all other existing methods, we focus on using downsized image and lightweight network to generate region proposals that only contain objects and speed up object detection process.
- To determine the optimal resolution, we investigated the relationship between resolution and accuracy.

Architecture

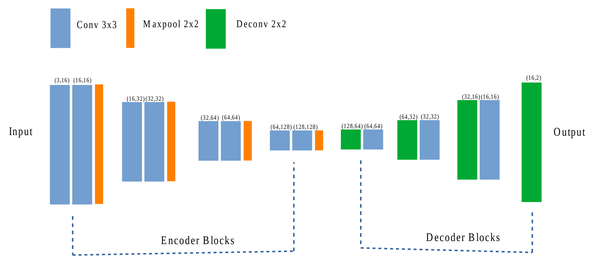

Figure 1:Proposed region proposal network.

Quantitative Results of Binary-class Detection

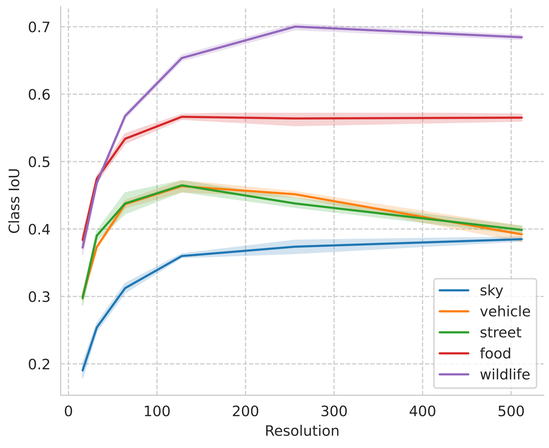

Figure 2:IoU in relation to input resolutions

Comparison of Speed and Computing Cost

Model	Backbone	Parameters(M)	FLOPs(G)	FPS
SSD	VGG16	26	31	82
RetinaNet	Resnet50	38	78	35
YOLOV4-tiny	Darknet53 tiny	6	2	238
YOLOV3	Darknet53	62	33	62
Faster R-CNN	Resnet50	28	52	14
UFO-RPN (256×256)	–	**0.38**	**1**	**988**
UFO-RPN (128×128)	–	**0.38**	**0.3**	**1011**

Table 1:Comparison with SOTA Models

Qualitative Results of Binary-class Detection

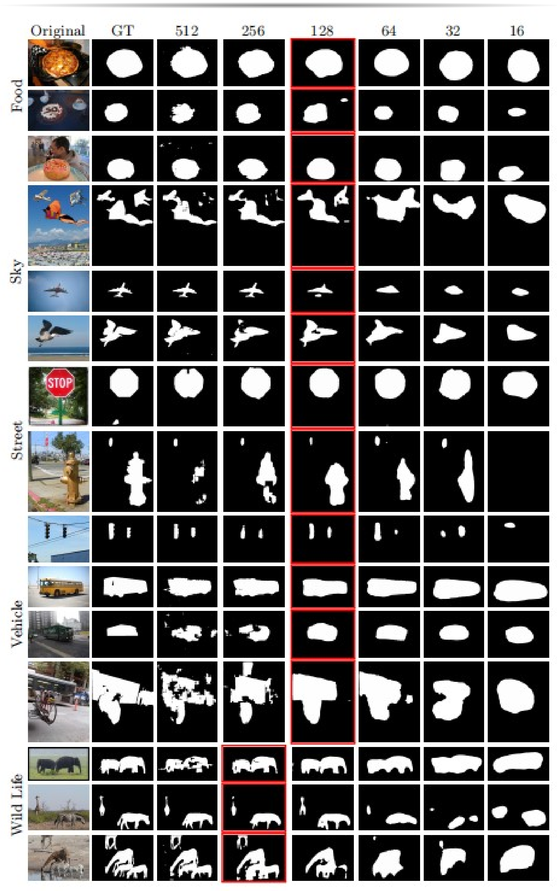

Figure 3:Detection output from different resolutions

FPS in relation to resolutions

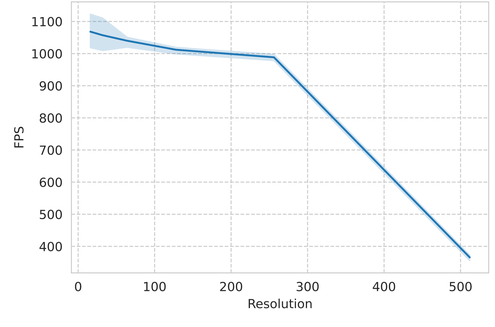

Figure 4:FPS in relation to resolutions

Data Set

- Five scenarios are taken out from MS COCO, including 'sky', 'street', 'vehicle', 'food', and 'wildlife'. Each scenario contains three classes of objects.
- Pixels of the three classes of objects in each scenario are all relabeled as '1', while the background pixels and non-target object pixels are relabeled as '0'.
- To explore the relationship between accuracy and resolution, we resized original images and labels to six variations: $\{16 \times 16, 32 \times 32, 64 \times 64, 128 \times 128, 256 \times 256, 512 \times 512\}$ with linear interpolation.

Conclusion

- Our proposed UFO RPN can greatly reduces computational cost by ignoring large proportion of non-valuable regions, e.g. background and non targets. Hence it can significantly improve inference speed without sacrificing detection accuracy.
- In terms of computing cost and inference speed, a comparison of FPS and FLOPs shows that our network significantly outperforms state-of-the-art approaches by two to three orders of magnitude.
- Our finding suggests that higher resolution is not necessarily the recipe for better accuracy. It is possible to achieve optimal performance at a lower resolution. The actual best resolution is likely task dependent and determined by the nature of the target object.

References

[1] Guo et.al, Fast object detection based on selective visual attention, Neurocomputing (2014)
[2] Yohanandan et.al, Saliency preservation in low-resolution grayscale images, ECCV (2018)

Contact Information

- Wenkai Li
- s3815738@student.rmit.edu.au