MULTI-SCALE OBJECT DETECTION WITH FEATURE FUSION AND REGION OBJECTNESS NETWORK

Wenjie Guan, Yuexian Zou, Xiaoqun Zhou

ADSPLAB/Intelligent Lab, School of ECE, Peking University, Shenzhen 518055, China

guanwenjie@pku.edu.cn, zouyx@pkusz.edu.cn

1 Introduction

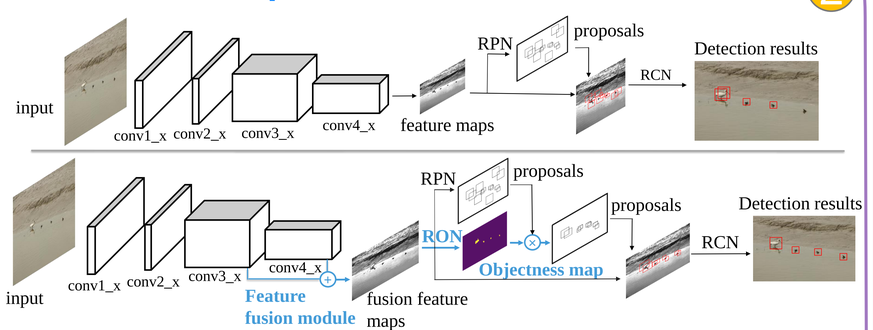
Recently, multi-scale object detection (MOD) draws considerable attention as it is highly demanded in real-world applications like costal wetland bird detection and vehicles detection for traffic surveillance. However, it meets bottleneck as follows:

- Mainstreams frameworks use high-level feature maps which is difficult to get the precise location of small objects.
- With many small objects for the MOD tasks, most of the object detection frameworks generate many redundant background proposals.

Proposed Solution

We proposed a new MOD method based on Faster R-CNN framework to tackle the problems mentioned above. Specifically, we introduce a feature fusion module to supplement the fine-grained knowledge for small objects in the final feature representation. Besides, a novel Region Objectness network is developed for generating effective proposals. In order to provide meaningful performance evaluation, experiments have been conducted over self-built costal wetland bird dataset (BSBDV 2017) and UA-DETRAC car dataset.

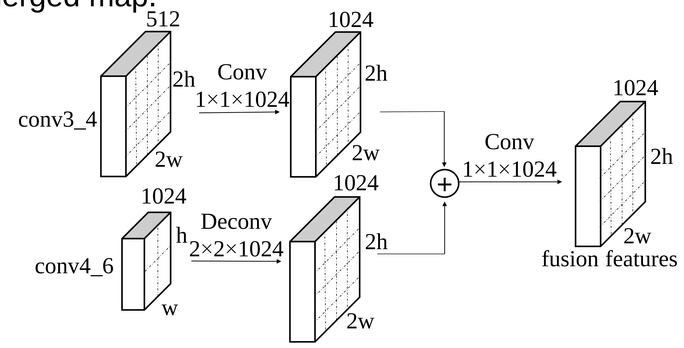
2 Proposed MOD Method



Different from Faster R-CNN, our MOD method newly added 2 parts, the feature fusion module and the Region Objectness Network (RON).We use ResNet-101 as the base network. To get better feature representation for small objects, feature fusion module outputs high-resolution feature maps. Besides, to eliminate the redundant background proposals, a novel binary objectness map generated by RON is proposed.

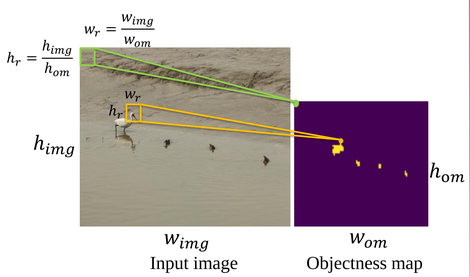
3 Feature Fusion Module

We choose conv3_4 and conv4_6 as the input of the feature fusion module, and outputs the finer resolution feature maps which contain the highly abstracted knowledge and fine-grained details of small objects. In order to further suppress the aliasing effect of the up-sampling process, a 1x1 conv layer is append on the merged map.



4 Region Objectness Network

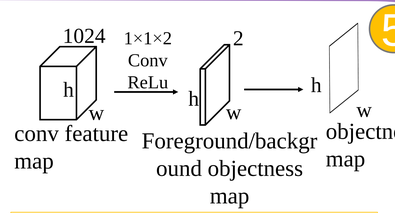
RON aims at eliminating the background proposals. We formulate the task as to predict the likelihood of each region in the input image being a foreground object as opposed to background. A RON takes an image of arbitrary size as input and outputs a binary objectness map. Each pixel of the objectness map only corresponds to a region in the image, which is called its *governing region* here. We model this process with a FCN. To generate the objectness map, we append a 1×1×2 conv layer after the last shared conv feature maps to learn the score which measures the likelihood of the corresponding *governing region* being a foreground object or a background one.



An example image and its corresponding objectness map. Each pixel of the objectness map corresponds to a governing region with fixed size in the image. Yellow pixels indicate foreground governing region (yellow box) and purple pixels indicate background governing region (green box). The size of the governing region is decided by the ratio of the input image size to the objectness map size.

5

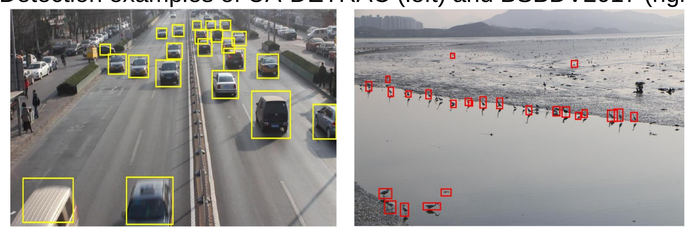
Our approach is supervised method. The loss function used is defined as formula (1). Where i is the governing region index, L_{cls} is the cross-entropy loss over two classes. p_i^* and N_{cls} represent ground-truth labels and the total number of governing regions respectively.



$$L(\{p_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) \quad (1)$$

6 Experiments

- Detection examples of UA-DETRAC (left) and BSBDV2017 (right)



- Detection results on UA-DETRAC (left) and BSBDV2017 (right)

Method	Base Network	AP (%)
YOLOv2	Darknet	44.3
SSD300	VGG-reduce	67
Faster R-CNN	ResNet-50	58.3
Faster R-CNN	ResNet-101	62.1
Ours	ResNet-101	**71.1**

Method	Base Network	AP (%)
YOLOv2	Darknet	34.6
SSD500	VGG-reduce	42
Faster R-CNN	ResNet-50	44.3
Faster R-CNN	ResNet-101	50.8
Ours	ResNet-101	**58.8**

- Detection speed on BSBDV 2017

Method	Base Network	AP (%)	Time (sec)	FPS
Faster R-CNN	ResNet-101	50.8	0.679	1.47
Ours	ResNet-101	**58.8**	**0.611**	**1.64**

7 Conclusions

- We proposed a multi-scale object detection method by introducing a feature fusion module and a novel Region Objectness Network, aiming at improving the localization performance of small objects and eliminating the redundant background proposals.
- To facilitate this study, a self-built bird dataset (BSBDV 2017) is established which will be made publicly available.