MASON: A Model AgnoStic ObjectNess Framework

K J Joseph[1] · Rajiv Chunilal Patel [2] · Vineeth N Balasubramanian [1]

[1]Indian Institute of Technology, Hyderabad, India, [2]ANURAG, DRDO, India

{cs17m18p100001,vineethnb}@iith.ac.in

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

GitHub link

1. Overview

- MASON is a simple and effective method to localize dominant foreground object(s) in an image to pixel level precision.
- The method uses pretrained image classification models to achieve the task, even without fine-tuning.
- MASON is:
  - ☐ Model agnostic: Works across different image classification models.
  - ☐ Category independent: Generalizes to a wide variety of classes.

2. MASON

- The activations obtained from the different layers of image classification networks like AlexNet, VGG-16 etc.; yield very good information for localizing the distinct object(s) in an image.
- We hypothesize that a linear weighted combination of the activations from such networks, helps to generate very useful heat map of the object(s) in an image.
- A simple sum of the activations from a deeper convolutional layer of such networks, yields the *objectness heatmap*, O, i.e:

$$O(I) = \sum_{i=1}^{|f^l|} f_i^l \tag{1}$$

where I refers to the input image, f^l refers to the feature maps at layer l and f_i^l refers to i^{th} feature map in layer l.
- The heatmap captures the likelihood of each pixel to contain an object.

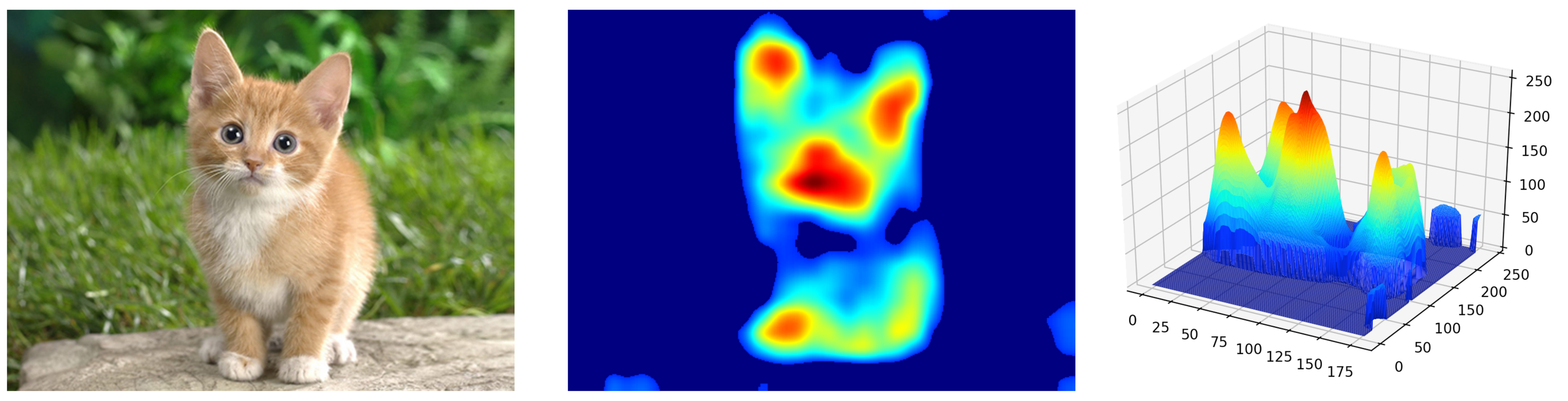

Figure 1. The figure shows the input image, the heatmap that MASON generates and its corresponding surface plot.

3. Fine-grained Object Localization using MASON

- The objectness heatmap $O(I)$ (Eqn 1) can be combined with GrabCut, an interactive graph cut based foreground extraction method, to generate fine-grained object localization in an image.
- Pixel intensities from $O(I)$ is stratified to automatically generate the foreground and background 'strokes' for GrabCut.
- This enables GrabCut to perform foreground extraction well, even when there are multiple objects in the scene, which is beyond the scope of the vanilla GrabCut algorithm.
- The method achieves 0.623532 mean IoU with the ground truth, when evaluated on Pascal VOC 2012 test set.

4. From Object Detection to Instance Segmentation using MASON

- Areas of the image inscribed by the bounding boxes predicted by an object detector can be passed through MASON + GrabCut method (explained in Section 3), to obtain foreground segmentation.
- This is combined with the class label information that the object detector predicts to generate instance segmentation.

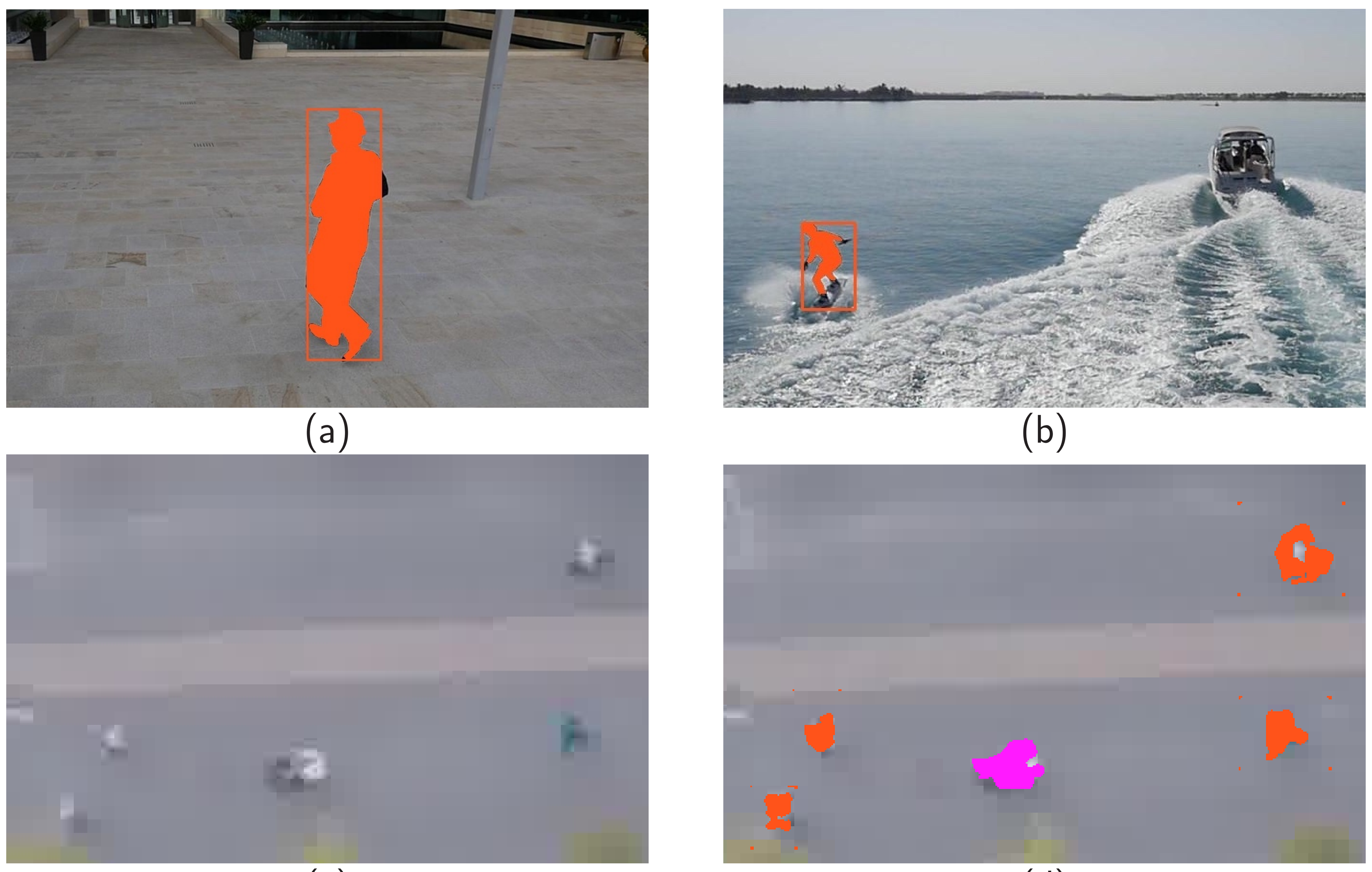

Figure 3. Instance Segmentation results from UAV123 Dataset and Stanford Drone Dataset.

5. Improving Quality of Detection Datasets using MASON

- MASON can provide an efficient and automated solution to:
  - ☐ filter out false positive annotations that actually contains no object inside its annotated area.
  - ☐ refine the true positive annotations, by making making the bounding box tighter around the object.
- The proposed method is especially useful for cleansing small object datasets, which are hard to annotate. Downstream applications like object detection can benefit from this automated preprocessing step as in illustrated in Table 1.

Class	Without Enhancement		With Enhanced Dataset	
	AP @ 0.5	AP @ 0.7	AP @ 0.5	AP @ 0.7
Pedestrian	0.302	0.093	0.763	0.542
Biker	0.303	0.124	0.711	0.521
Skater	0.144	0.092	0.589	0.436
Car	0.330	0.228	0.554	0.440
Bus	0.241	0.282	0.591	0.553
Cart	0.577	0.420	0.750	0.684
mAP	0.316	0.207	0.659	0.529

Table 1. Table shows the mAP values of object detection on Stanford Drone Dataset (SDD) using R-FCN object detector, with and without using enhanced annotations.

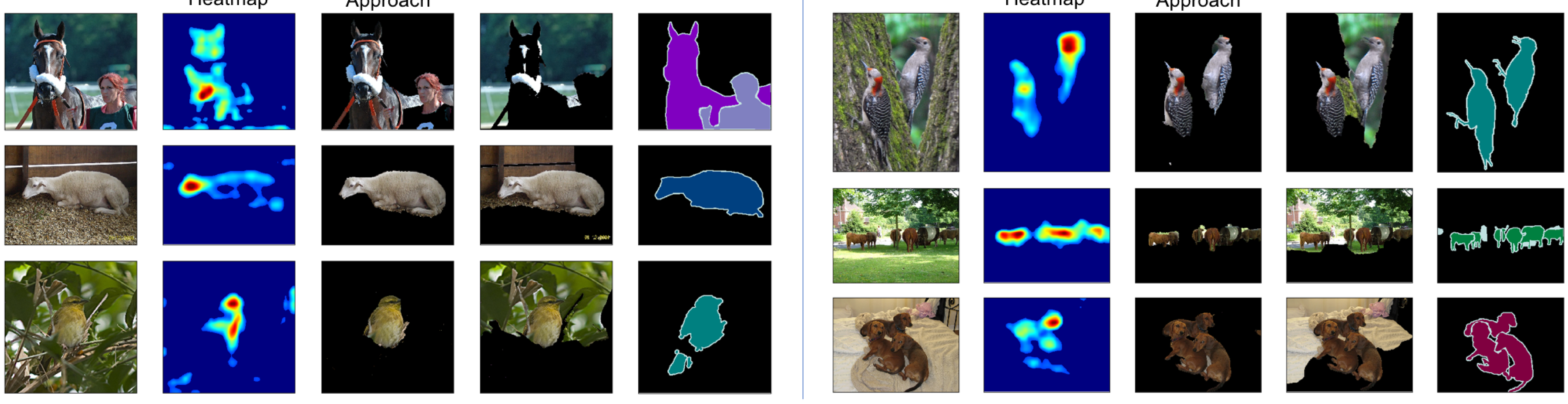

Figure 2. The effectiveness of MASON in fine-grained object localization is captured in the figure (Column 3). The proposed method (Column 3) is more effective than just using Grabcut (Column 4) in foreground segmentation.

First International Workshop On Autonomous Navigation in Unconstrained Environments, European Conference on Computer Vision(ECCV), Munich, Germany, September 2018.