

# Attend Refine Repeat: Active Box Proposal Generation via In-Out Localization

Spyros Gidaris  
 spyros.gidaris@enpc.fr  
 Nikos Komodakis  
 nikos.komodakis@enpc.fr

Université Paris-Est, École des Ponts  
 ParisTech  
 Paris, France

**Introduction:** In our work we deal with the problem of *category agnostic box proposal generation*. Its definition is that for a given image a small set of boxes must be generated that will cover with high recall all the objects in the image regardless of their category. Recently, this problem has received an immense amount of attention due to the fact that box proposal generators have become a core component in many vision tasks, ranging from object detection till visual question answering, leading in all of them to state-of-the-art results.

**Approach:** The dominant paradigm in box proposal generation is that of having a CNN model that given a set of input boxes (uniformly distributed in the image), it predicts their objectness score and refines their coordinates such that they better align with object’s borders (i.e. bounding box prediction). In that context, our work improves the previous state-of-the-art in box proposal generation in two ways. (1) We improve the object’s bounding box prediction step by adapting the successful LocNet [1] approach for category-specific object localization to that of category-agnostic localization (see Figure 1). (2) We employ an active box proposal generation strategy, which we call *Attend Refine Repeat* algorithm (see algorithm on the right), that starts from a set of seed boxes, which only depend on the image size, and it then sequentially produces newer boxes that will better cover the objects of the image while avoiding the "objectless" image areas (see Figure 2).

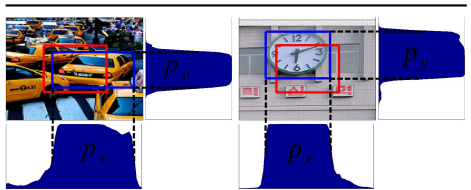
**Results:** We exhaustively evaluate our system both on PASCAL and on the more challenging COCO datasets and we demonstrate significant improvement with respect to the state-of-the-art on box proposal generation. Furthermore, we provide strong evidence that our object location refinement module is capable of generalizing to unseen categories.

[1] Spyros Gidaris and Nikos Komodakis. Locnet: Improving localization accuracy for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

## Algorithm: Attend Refine Repeat

```

Input : Image I
Output: Bounding box proposals P
C ← ∅, B0 ← seed boxes
for i ← 1 to T do
    /* Attend & Refine procedure */
    O' ← ObjectnessScoring(Bi-1|I)
    B' ← ObjectLocationRefinement(Bi-1|I)
    C ← C ∪ {B', O'}
end
P ← NonMaxSuppression(C)
    
```



**Figure 1: Object location refinement formulation.** We formulate the problem of box prediction to that of dense classification. Specifically, given an input box (red rectangle), our model defines a search region (depicted image crop) and assigns a membership probability to each row and each column of that region ( $p_y$  and  $p_x$  probability vectors) that represent the likelihood of those elements (rows or columns) to be inside the target bounding box (blue rectangle).



**Figure 2: Image areas that are attended by our active proposal generation algorithm as it progresses from the first (left column) to the last iteration (right column).**

Method	AR@10	AR@100	AR@1000
EdgeBoxes	0.074	0.178	0.338
Geodesic	0.040	0.180	0.359
Selective Search	0.052	0.163	0.357
MCG	0.101	0.246	0.398
DeepMask	0.153	0.326	0.482
Co-Obj	0.189	0.366	0.492
SharpMask	0.192	0.391	0.555
AttractionNet (Ours)	<b>0.328</b>	<b>0.535</b>	<b>0.661</b>

**Table 1: Sampled average recall results on COCO validation set.**