How good are detection proposals, really?

Jan Hosang
http://mpi-inf.mpg.de/~jhosang
Rodrigo Benenson
http://mpi-inf.mpg.de/~benenson
Bernt Schiele
http://mpi-inf.mpg.de/~schiele

MPI Informatics
Saarbrücken, Germany

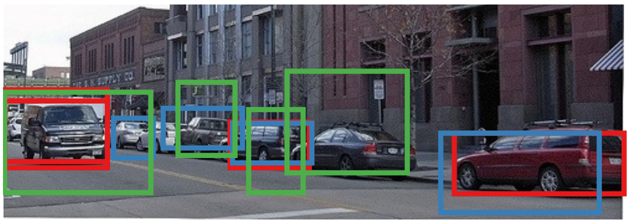

Figure 1: Each colour indicates a different set of detection proposals. How to evaluate which method provides the best proposals?

Object detection is traditionally instantiated in the well known "sliding window" paradigm where a classifier is evaluated over an exhaustive list of positions, scales, and aspect ratios. This approach evaluates the classifier at about $\sim 10^6$ different locations. To alleviate computation pressure by avoiding exhaustive search, recent detectors delegate the selection of candidate detections to a pre-processing step. If these class-agnostic candidate detectors can achieve high recall with $\sim 10^4$ or less windows, significant speed-ups can be achieved, enabling the use of more sophisticated classifiers.

Current top performing Pascal VOC object detectors employ detection proposals to guide the search for objects, thereby avoiding exhaustive sliding window search across images. Despite the popularity of detection proposals, it is unclear which trade-offs are made when using them during object detection. We provide an in depth analysis of ten object proposal methods (from 2009 to 2014) along with four baselines regarding: a) ground truth annotation recall (on Pascal VOC 2007 and ImageNet 2013), b) repeatability, and c) impact on DPM detector performance. See table 1. Our findings show common weaknesses of existing methods, and provide insights for practitioners seeking to choose the most adequate method for their application.

Repeatability We introduce the notion of repeatability which captures how much a detection proposal method is affected by different image perturbations. For this analysis we compute how well a method repeats the selection of candidates after applying an image transformation (see figure 2 for perturbation examples). We argue that repeatability is important when a detector uses candidate detection for negative mining, as it requires the distribution of negative windows to be very similar between training and test set. Our results indicate that repeatability seems to be an issue for most methods. Even very small changes cause most methods to have a strong drop in repeatability.

Recall Different papers evaluate based on recall at different operating points. We give a full picture evaluation regarding recall of ground-truth bounding boxes, and establish common ground for a proper comparison between different methods. To this end we analyse the recall as a function of both the number of candidates and the localisation quality. Recall is important because objects lost by the proposal method will not be recovered by the detector. Our results show that a handful of methods dominate quality in multiple settings (see figure 3). The ImageNet experiments show that, despite being tuned on Pascal VOC, current proposal methods have excellent generalization towards the larger (10×) set of ImageNet classes, indicating that they are true "objectness" measures.

Detection Finally, we do experiments regarding the effect of selective search over detection quality. As an initial approach, we filter the detections of a pre-trained DPM detector method using different proposal methods, to emulate having done the detections directly from these windows. Results show that detection quality is directly related to the accuracy and recall level of the underlying detection proposals method.

Our paper provide detailed result curves and tables, summarising more than 500 experiments over different data sets, totalling to more than 2.5 months of CPU computation.

Figure 2: Example of the image perturbations considered for the repeatability experiments. Top to bottom, left to right: original, then blur, illumination, JPEG artefact, rotation, and scale perturbations.
How stable are detection proposals to slight changes in the input image?

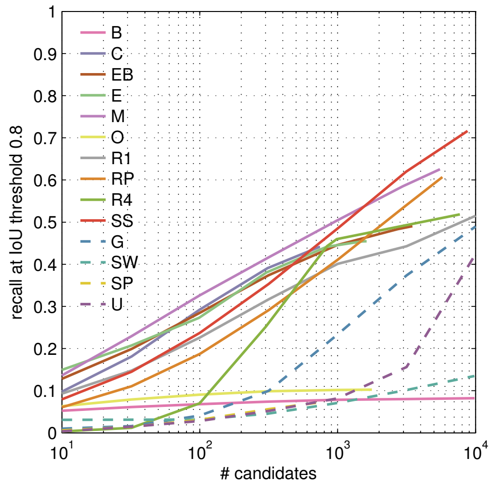

Figure 3: Recall versus number of proposals on the Pascal VOC 2007 test set for IoU above 0.8. At varying number of windows the best method to pick changes, what about when considering computational cost too?

Method		Time	Repeatability	Recall	Detection
Objectness	O	3	·	⋆	·
CPMC	C	250	-	⋆⋆	⋆
Endres2010	E	100	-	⋆⋆	⋆⋆
Sel.Search	SS	10	⋆⋆	⋆⋆⋆	⋆⋆
Rahtu2011	R1	3	·	·	⋆
Rand.Prim	RP	1	⋆	⋆	⋆
Bing	B	0.2	⋆⋆⋆	⋆	·
MCG	M	30	⋆	⋆⋆⋆	⋆⋆
Ranta.2014	R4	10	⋆⋆	·	⋆
EdgeBoxes	EB	0.3	⋆⋆	⋆⋆⋆	⋆⋆
Uniform	U	0	·	·	·
Gaussian	G	0	·	·	⋆
SlidingWindow	SW	0	⋆⋆⋆	·	·
Superpixels	SP	1	⋆	·	·

Table 1: Overview of detection proposal methods.
Time is in seconds. Repeatability, quality, and detection rankings are provided as rough qualitative overview; "-" indicates no data, "·", "⋆", "⋆⋆", "⋆⋆⋆" indicate progressively better results. See paper's text for details and quantitative evaluations.