CAP 6412 Advanced Computer Vision

http://www.cs.ucf.edu/~bgong/CAP6412.html

Boqing Gong

Jan 26, 2016

Today

- Administrivia
- A bigger picture and some common questions
- Object detection proposals, by Samer

Past due (12pm today)

- Assignment 2: Review the following paper

{**Major**} [Detection Proposals] J. Hosang, R. Benenson, P. Dollár, and B. Schiele. What makes for effective detection proposals? PAMI 2015.

Template for paper review:
http://www.cs.ucf.edu/~bgong/CAP6412/Review.docx

An assignment with no due dates

- See “Paper Presentation” on UCF webcourse

- Sharing your slides
  - **Refer to the originals sources of images, figures, etc. in your slides**
  - Convert them to a PDF file
  - Upload the PDF file to “Paper Presentation” after your presentation

Schedule update

Week 2	CNN visualization & object recognition
Week 3	CNN & object localization
Week 4	**CNN & transfer learning**
Week 5	CNN & segmentation, super-resolution
Week 6	CNN & videos (optical flow, pose)
Week 7	**Image captioning & attention model**
Week 8	Visual question answering
Week 9	Attention model, aligning books with movies
Week 10--16	Video: tracking, action, surveillance Human-centered CV 3D CV Low-level CV, etc.

Next week:
Image captioning & attention model

Tuesday (02/02) Harish Ravi Prakash	Karpathy, Andrej, and Li Fei-Fei. “Deep visual-semantic alignments for generating image descriptions.” arXiv preprint arXiv:1412.2306 (2014). & Secondary papers
Thursday (02/04) Karan Daei-Mojdehi	Xu, Kelvin, Jimmy Ba, Ryan Kiros, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, and Yoshua Bengio. “Show, attend and tell: Neural image caption generation with visual attention.” arXiv preprint arXiv:1502.03044 (2015). & Secondary papers

Beginning next class

- Make good presentations --- #3 course objective
- Title, authors (full name), authors' institutes, your name and email
- Motivation of the research (1—2 slides)
- Problem statement (1—2 slides)
- Main contributions of the paper
- Approach outline (1 slide)
- Details of the proposed approach
- Experiments
- Related work (1—3 slides)
- Conclusion: take-home message (1—2 slides)
- Strengths & weaknesses of the paper (1—2 slides)
- Overall rating & why (how you weigh the strengths and weaknesses) (1 slide)
- Future directions (1—3 slides)

Beginning next class

- Make good presentations --- #3 course objective
- Title, authors (full name), authors' institutes, your name and email
- Motivation of the research (1—2 slides)
- Problem statement (1—2 slides)
- Main contributions of the paper
- Approach outline (1 slide)
- Details of the proposed approach
- Experiments
- Related work (1—3 slides)
- Conclusion: take-home message (1—2 slides)
- Strengths & weaknesses of the paper (1—2 slides)
- Overall rating & why (how you weigh the strengths and weaknesses) (1 slide)
- Future directions (1—3 slides)

40 mins only

Leave me time to cover:

- Underexploited points in slides/discussion
- Technique details
- More related work and reading references
- My own comments

Today

- Administrivia
- A bigger picture and some common questions
- Object detection proposals, by Samer

Why we read these papers: A personalized and biased perspective

Why we read these papers: A personalized and biased perspective

Time	Event	Related Papers	Read?
01/2012	Negative CVPR reviews	[LeNet] **Yann LeCun**, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, november 1998.	Yes

Why we read these papers: A personalized and biased perspective

Time	Event	Related Papers	Read?
01/2012	Negative CVPR reviews	[LeNet] **Yann LeCun**, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, november 1998.	Yes
10/2012	AlexNet wins ILSVRC2012	**[AlexNet]** Krizhevsky, Alex, Ilya Sutskever, and **Geoffrey E. Hinton.** "Imagenet classification with deep convolutional neural networks." In NIPS, 2012.	Yes

Why we read these papers: A personalized and biased perspective

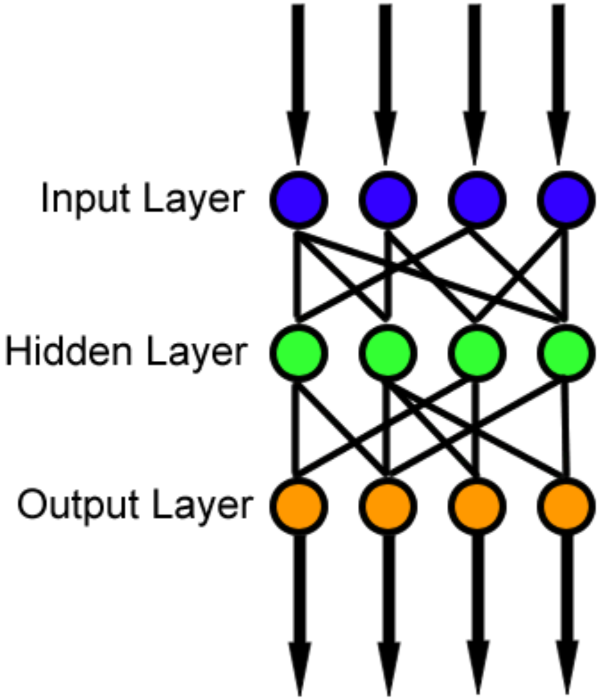
Time	Event	Related Papers	Read?
01/2012	Negative CVPR reviews	[LeNet] **Yann LeCun**, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, november 1998.	Yes
10/2012	AlexNet wins ILSVRC2012	**[AlexNet]** Krizhevsky, Alex, Ilya Sutskever, and **Geoffrey E. Hinton.** "Imagenet classification with deep convolutional neural networks." In NIPS, 2012.	Yes
11/2013	Visualize & understand CNNs	**[Visualization]** Zeiler, Matthew D., and **Rob Fergus**. "Visualizing and understanding convolutional networks." In ECCV, 2014.	Yes

Why we read these papers: A personalized and biased perspective

Time	Event	Related Papers	Read?
01/2012	Negative CVPR reviews	[LeNet] **Yann LeCun**, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, november 1998.	Yes
10/2012	AlexNet wins ILSVRC2012	**[AlexNet]** Krizhevsky, Alex, Ilya Sutskever, and **Geoffrey E. Hinton.** “Imagenet classification with deep convolutional neural networks.” In NIPS, 2012.	Yes
11/2013	Visualize & understand CNNs	**[Visualization]** Zeiler, Matthew D., and **Rob Fergus**. “Visualizing and understanding convolutional networks.” In ECCV, 2014.	Yes
2014	CNN wins on object detection	**Girshick, Ross**, Jeff Donahue, Trevor Darrell, and Jagannath Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." In *CVPR*, 2014.	This Thursday

Basic network structures --- where is CNN?

- Feed-forward networks



- Recurrent neural networks

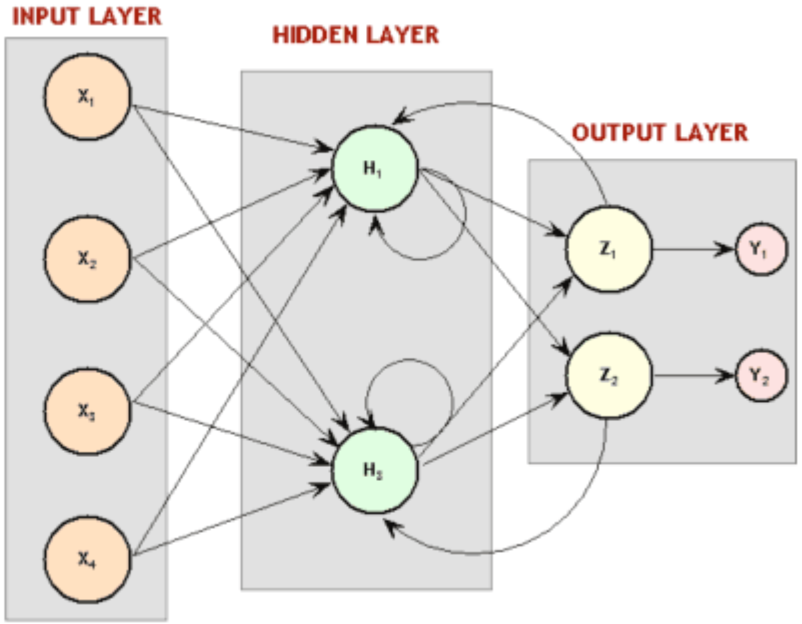
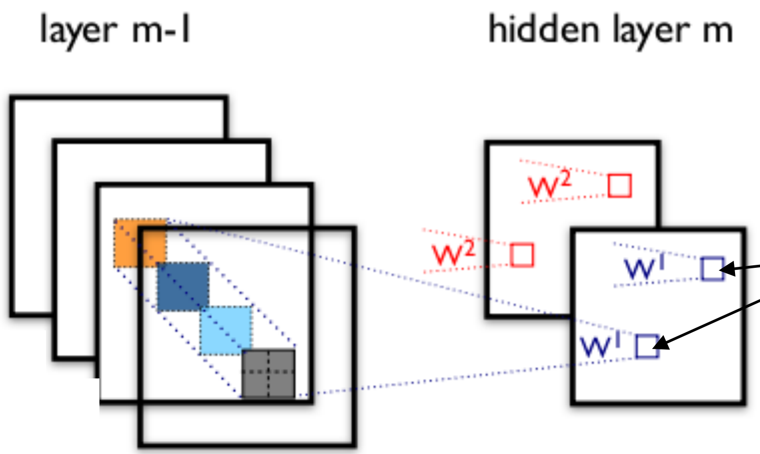


Image credit: http://mesin-belajar.blogspot.com/2016/01/a-brief-history-of-neural-nets-and-deep_84.html

CNN: a special form of feed-forward networks

- See whiteboard

Detour: Weight sharing in CNN



Convolution layer

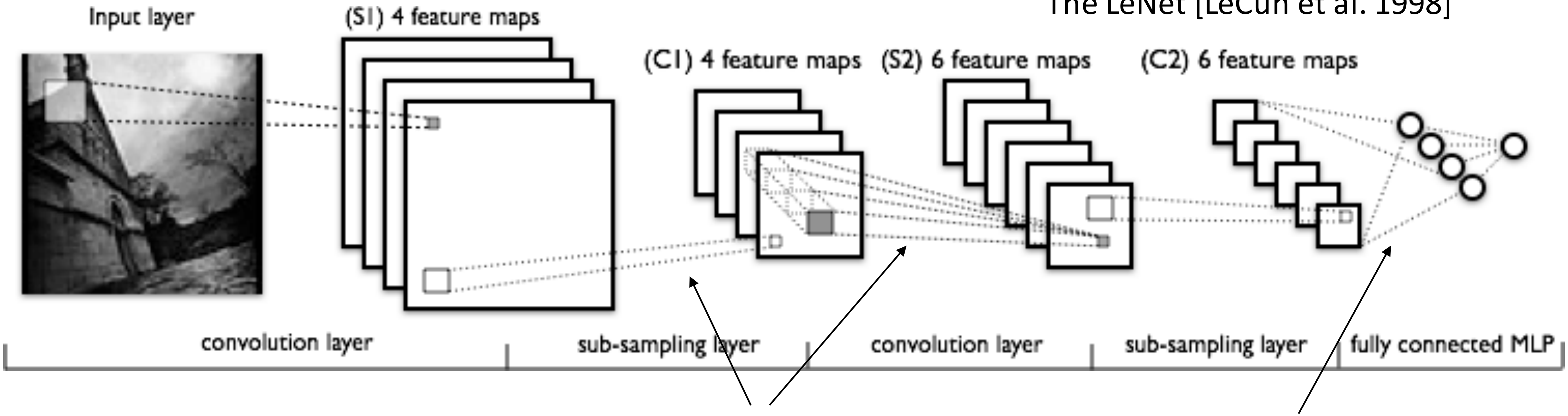
Neurons of the same feature map share the same weights (the filter)

Significantly reduced # parameters

Image credit: deeplearning.net/tutorial/lenet.html

Detour: Sparse connection in CNN

The LeNet [LeCun et al.'1998]



Sparse connections *vs.* Full connection

Smaller # parameters,
better learning efficiency

Today

- Administrivia
- A bigger picture and some common questions
- Object detection proposals, by Samer

What makes for effective detection proposals?

Jan Hosang[1], Rodrigo Beneson[1], Piotr Dollar[2], and Bernt Schiele[1]

[1]Max Planck Institute for Informatics

[2]Facebook AI Research (FAIR)

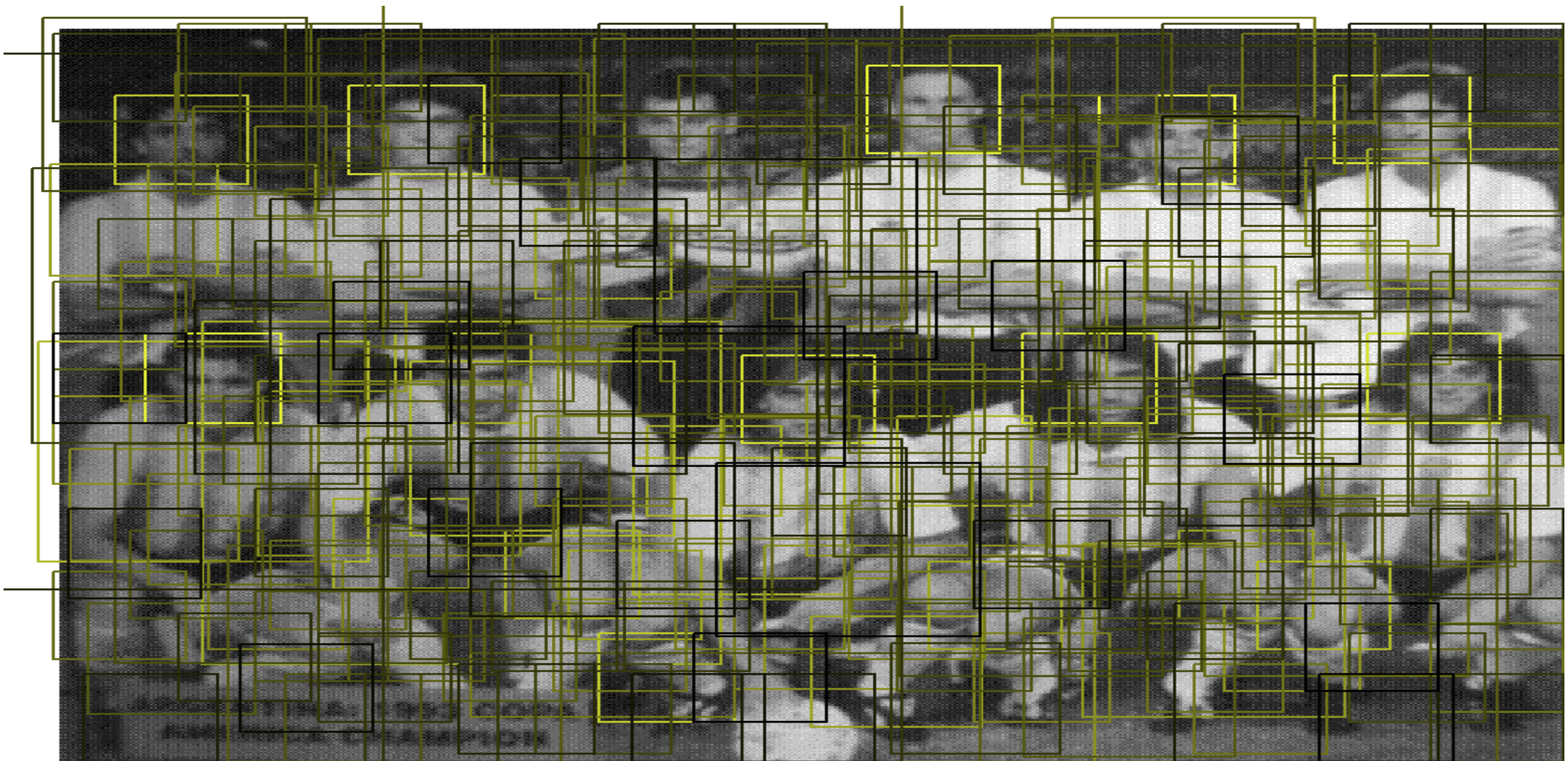
Presented by:

Samer Iskander

(samers.iskander@knights.ucf.edu)

Motivation

- High performing object detectors are based on object proposals, in order to avoid exhaustive sliding window search across the image.



- As a result of that, an in-depth analysis of different methods is required, in order to study their impact on detection performance.

Problem Statement

- Although the widespread use of detection proposals, it is necessary to study the performance metrics trade-offs when employing them.

Main Contributions

- A systematic overview of detection proposal methods is provided.
- The notion of proposal repeatability is introduced.
- Object recall metric is studied on different datasets.
- The influence of different proposal methods when applied on selected objects detection algorithms (DPM, R-CNN and Fast R-CNN).
- A novel metric, the average recall (AR), which rewards both proposal localization and recall performance metrics and effects the detection performance is proposed.

Approach Outline

1. Detection Proposal Methods
   1.1 Baseline Proposal Method
2. Evaluation Metrics for Object Proposals
3. Proposal Repeatability
4. Proposal Recall
5. Using The Detection Proposals
   5.1 Detector Responses Around Objects
   5.2 LM-LLDA, R-CNN and Fast R-CNN detection performance
   5.3 Predicting detection performance

Details of The Proposed Approach

1. Detection Proposal Methods

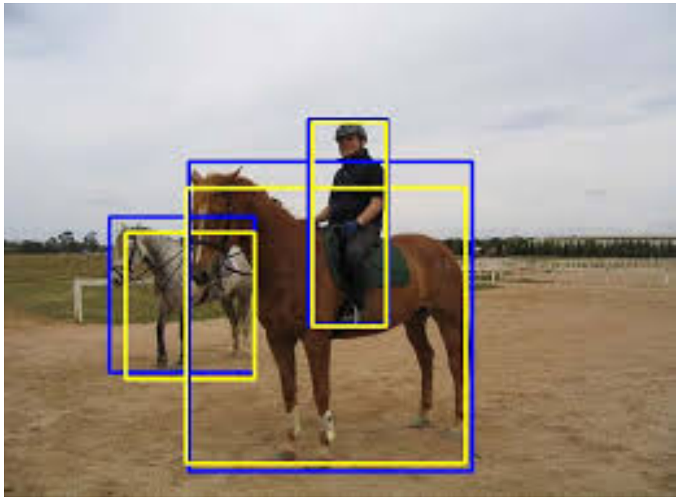
Detection Proposal Methods

Grouping Proposal Methods

- They attempt to generate segments (may be overlapped) that are likely to correspond to objects

Window Scoring Methods

- They score each candidate window according to how likely it is to contain an object.
- It is faster.
- If not generates densely windows, low localization accuracy

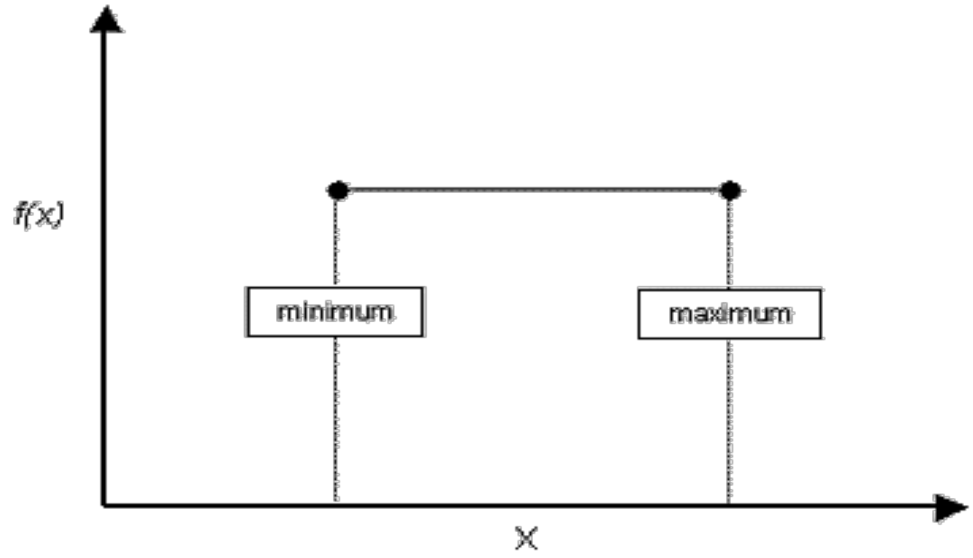


1.1 Baseline Proposal Method

A. Uniform:

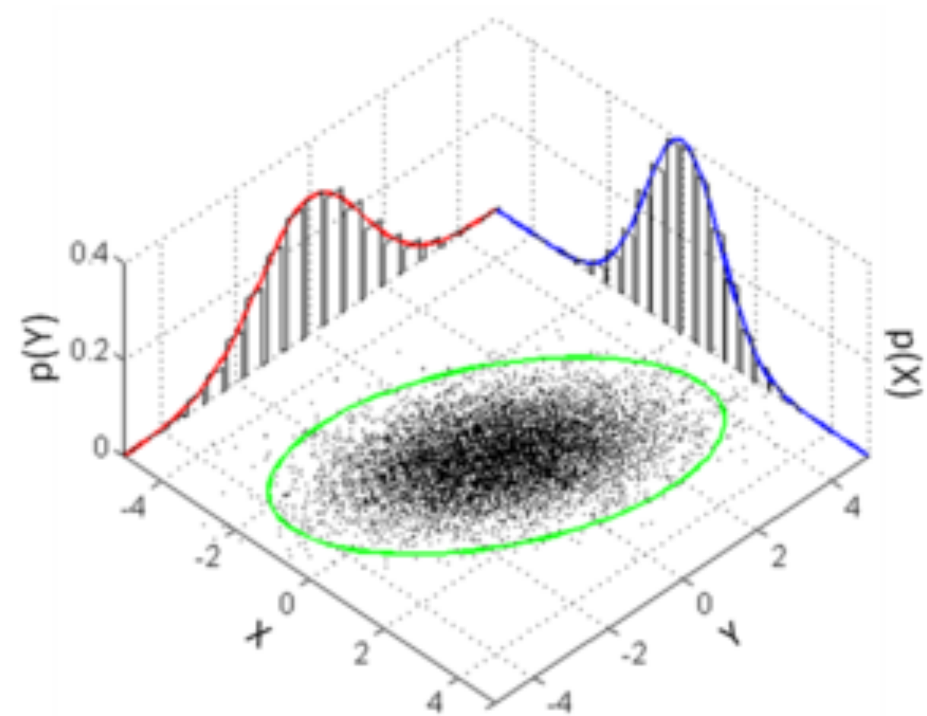
To generate proposals, it is necessary to uniformly sample the bounding box center position (x,y), square root area and log aspect ratio.

The PASCAL VOC 2007 training set is used to estimate these parameters.



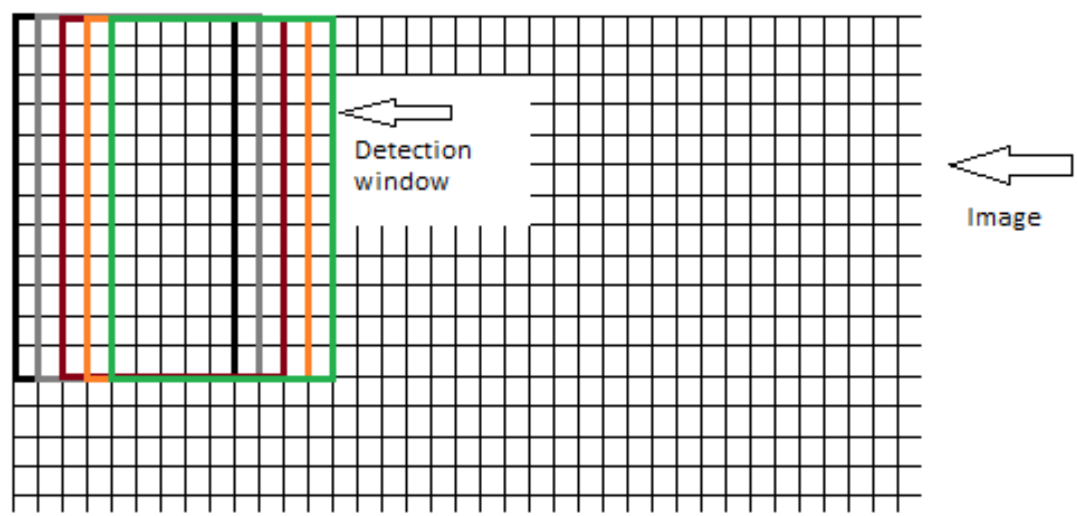
B. Gaussian:

To generate proposals, it is necessary to multivariate Gaussian distribution the bounding box center position (x,y), square root area and log aspect ratio.



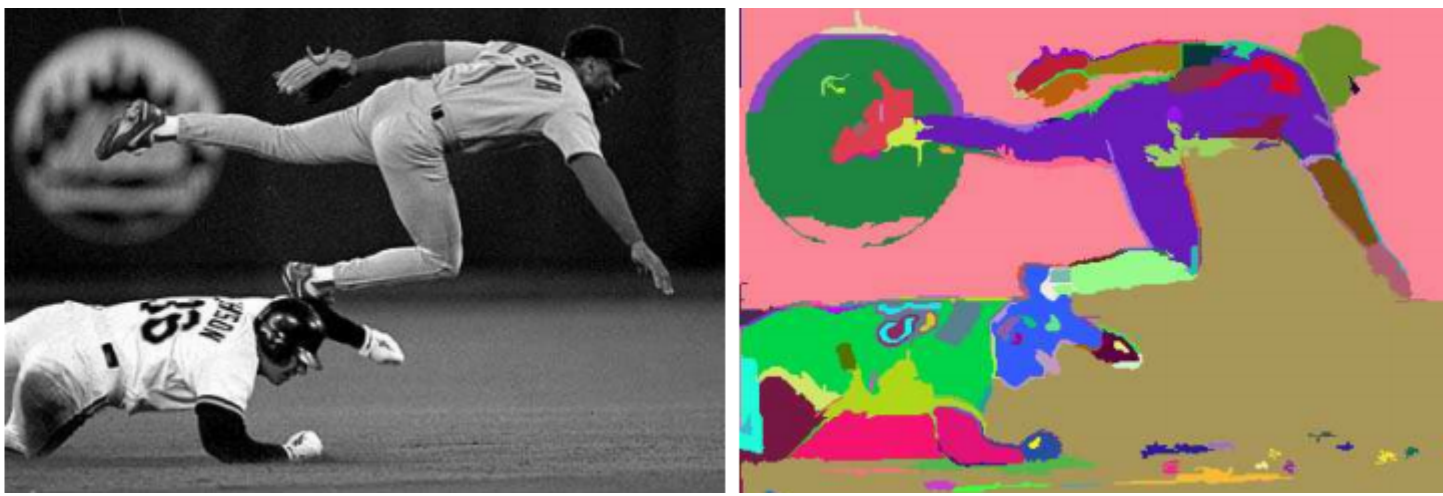
C. Sliding Window:

Equally distributed windows in space are generated. BING (Binarized Normed Gradients for Objectness Estimation at 300fps) uses 29 specific sizes, this method spread this sizes homogeneously inside the image.



D. Superpixels:

Superpixels are generated from Efficient Graph-Based Image Segmentation.

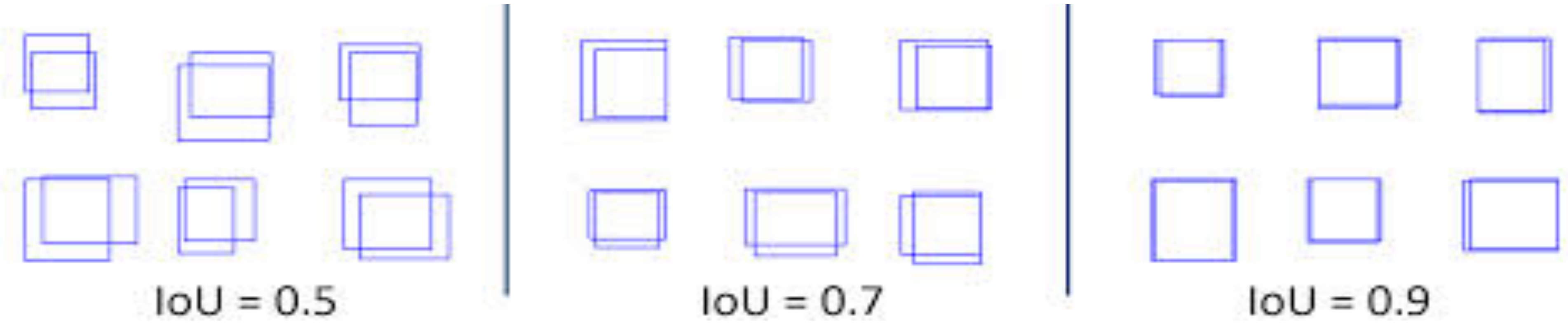


2. Evaluation Metrics for Object Proposals

1. Intersection Over Union (IOU):

- The metrics used for evaluating object proposals are all typically functions of intersection over union (IOU) between generated proposals and ground-truth annotations.
- For two boxes/regions b_i and b_j , IOU is defined as:

$$IOU(b_i, b_j) = \frac{area(b_i \cap b_j)}{area(b_i \cup b_j)}$$



2. Recall @ IOU Threshold t:

- For each ground-truth instance, check whether the best proposal from list L has IOU > t.
- If so, this ground-truth instance is considered detected or recalled.
- Then average recall is measured overall the ground-truth instances.

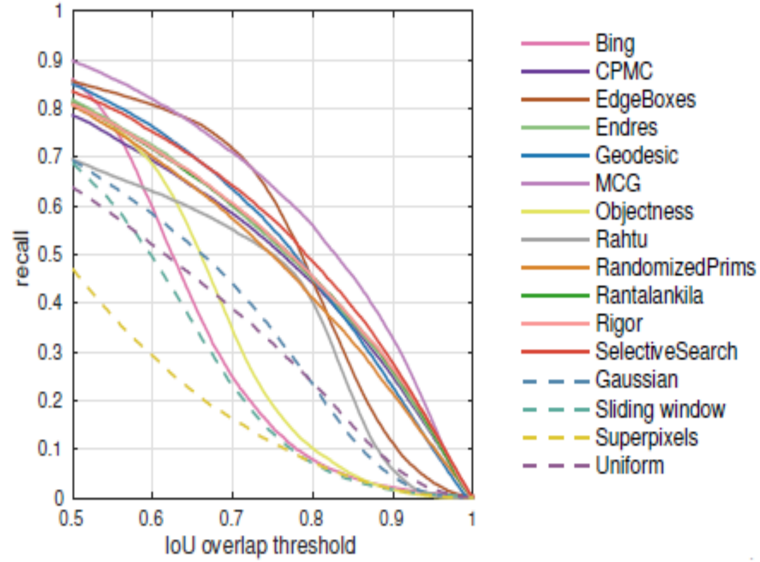
$$recall@t = \frac{1}{|G|} \sum_{g_i \in G} I\left[\max_{l_i \in L} IOU(g_i, l_i) > t\right]$$

I[.] is an indicator function for logical preposition in the argument

- Object proposals are evaluated using this metric in two ways:

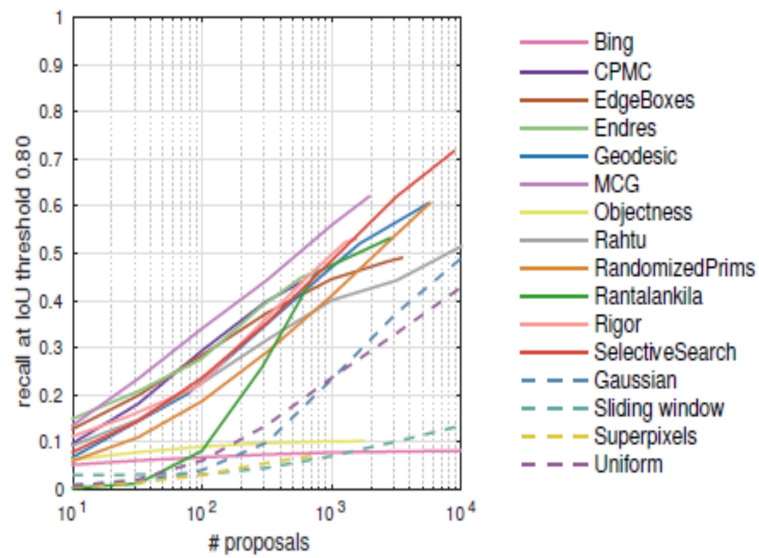
1. Plotting recall vs. t by fixing #proposals in L.

2. Plotting recall vs. # proposals by fixing t.



1 000 proposals per image.

Recall versus IoU threshold on the PASCAL VOC 2007 test set.



Recall at 0.8 IoU.

Recall versus number of proposal windows on the PASCAL VOC 2007 test set.

3. Average Best Overlap (ABO):

This metric eliminates the need for the threshold.

Calculate the overlap between each ground-truth annotation $g_i \in G$ and the best object hypothesis in L.

$$ABO = \frac{1}{|G|} \sum_{g_i \in G} \max_{l_i \in L} IOU(g_i, l_i)$$

4. Average Recall (AR):

$$ABO = \frac{1}{|G|} \sum_{g_i \in G} \max_{l_i \in L}(IOU(g_i, l_i)\text{-}0.5,0)$$

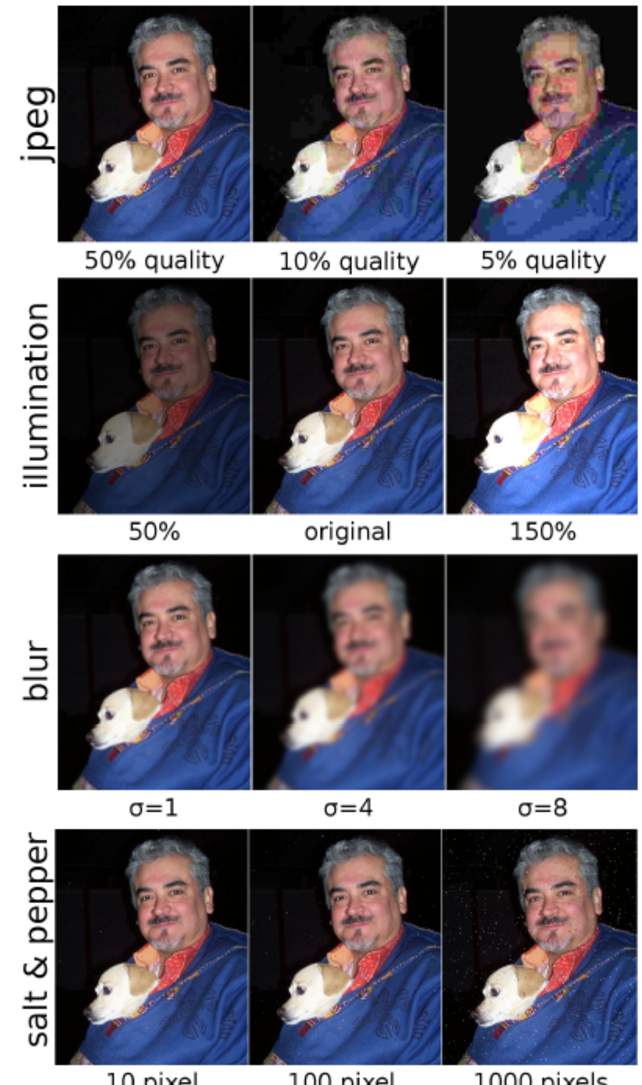
Average recall (for IOU between 0.5:1)vs. #proposals

5. Volume Under Surface (VUS):

It plots recall as a function of both t and #proposals and computes the volume under the surface.

3. Proposal Repeatability

1. For each image in the PASCAL VOC 2007 test set, several perturbed versions are generated (blur, rotation, scale, illumination, JPEG compression, and "salt and pepper" noise).



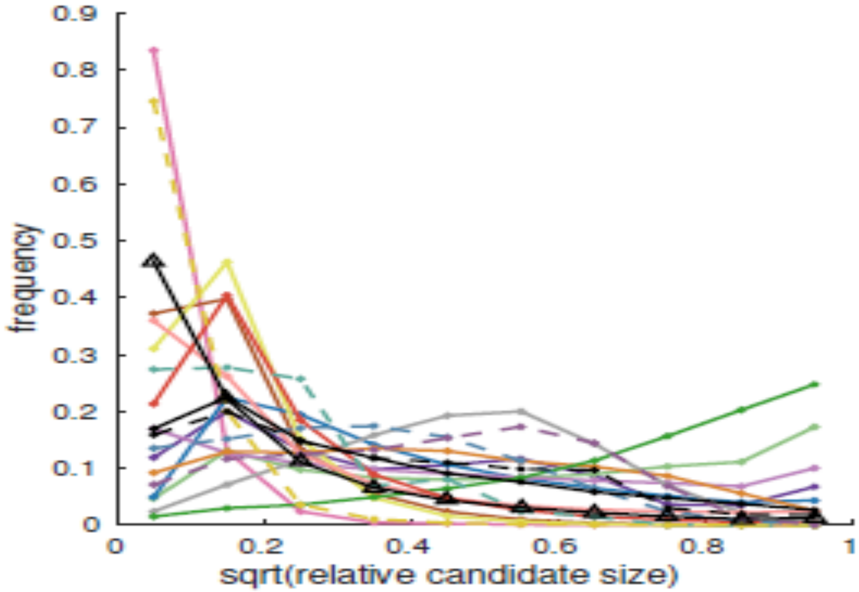
2. For each pair of reference and perturbed images, detection proposals are computed with a given method (generating 1000 windows per image).

3. The proposals are projected back from the perturbed into the reference image and then matched to the proposals in the reference image.

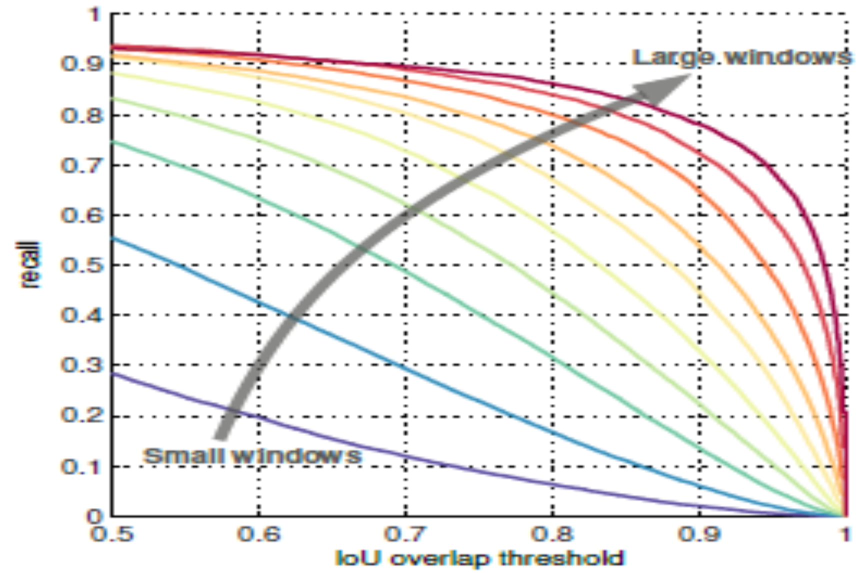
4. Then, plot recall vs. IOU t (0:1), and repeatability is the area under the curve.

5. Methods that propose windows at similar locations at high IoU—and thus on similar image content—are more repeatable, since the area under the curve is larger.

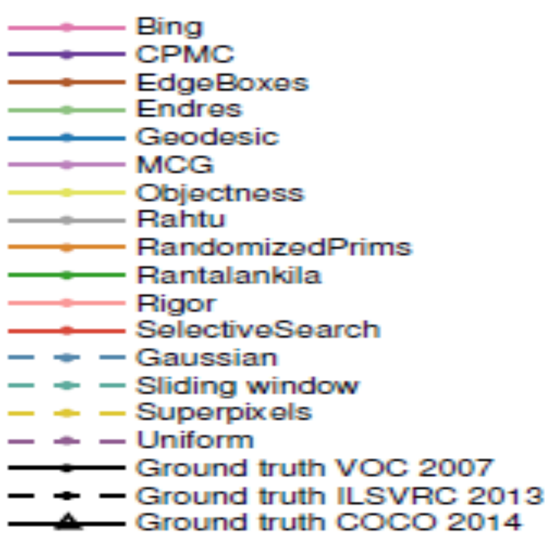
6. Large windows are more likely to match than smaller ones since the same perturbation will have a larger relative effect on smaller windows.



(a) Histogram of proposal sizes on PASCAL.



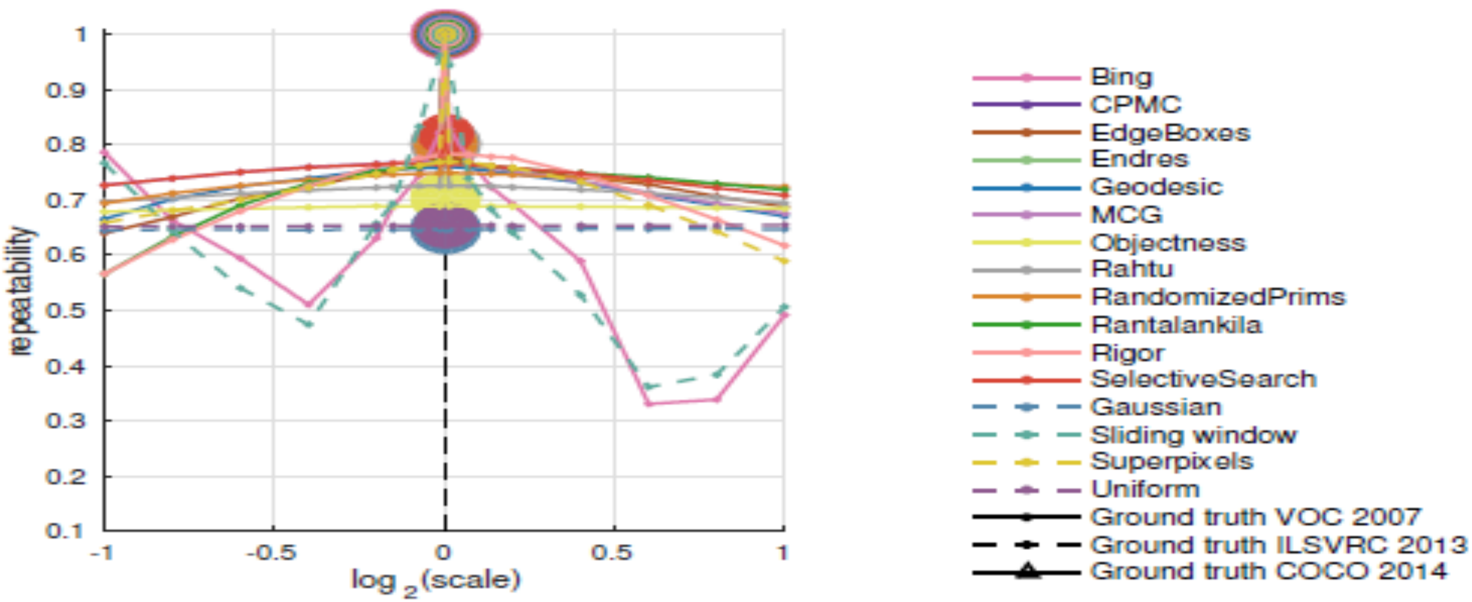
(b) Example of recall at different scales.



- **<u>Scale:</u>**

All methods except Bing show a drastic drop with small scale changes, but suffer only minor degradation for larger changes.

Bing is more robust to small scale changes; however, it is more sensitive to larger changes due to its use of a coarse set of box sizes while searching for candidates.

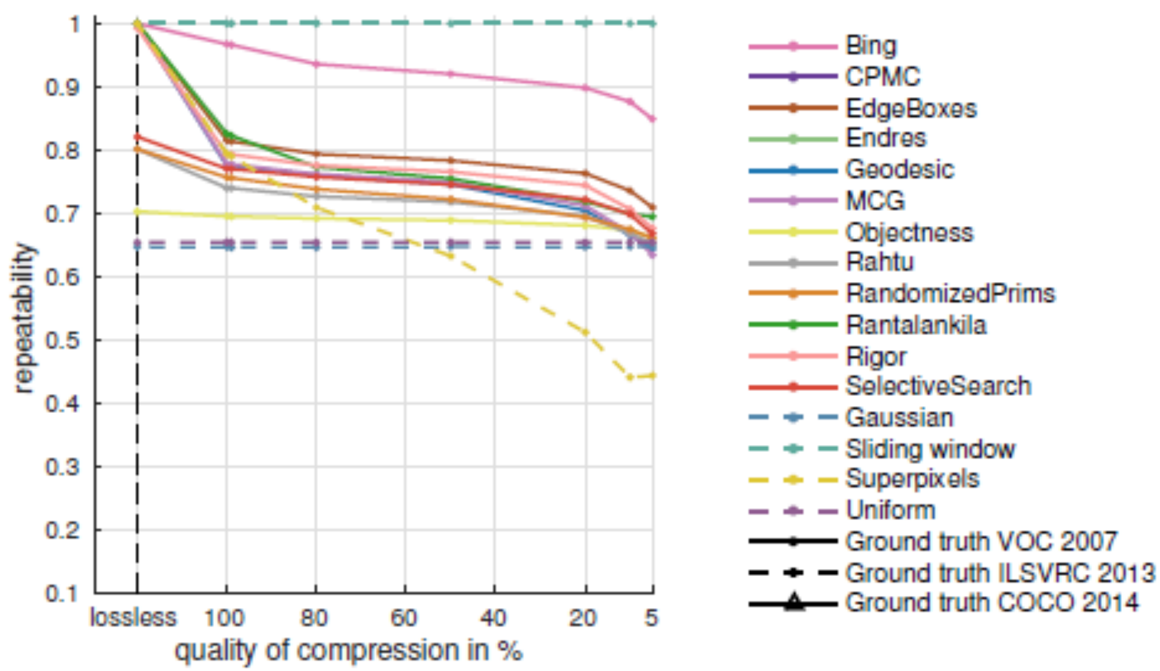


Scale.

- **JPEG Compression:**

Small compression has a large effect and more aggressive compression shows monotonic degradation.

Despite using gradient information, Bing is most robust to these kind of changes.

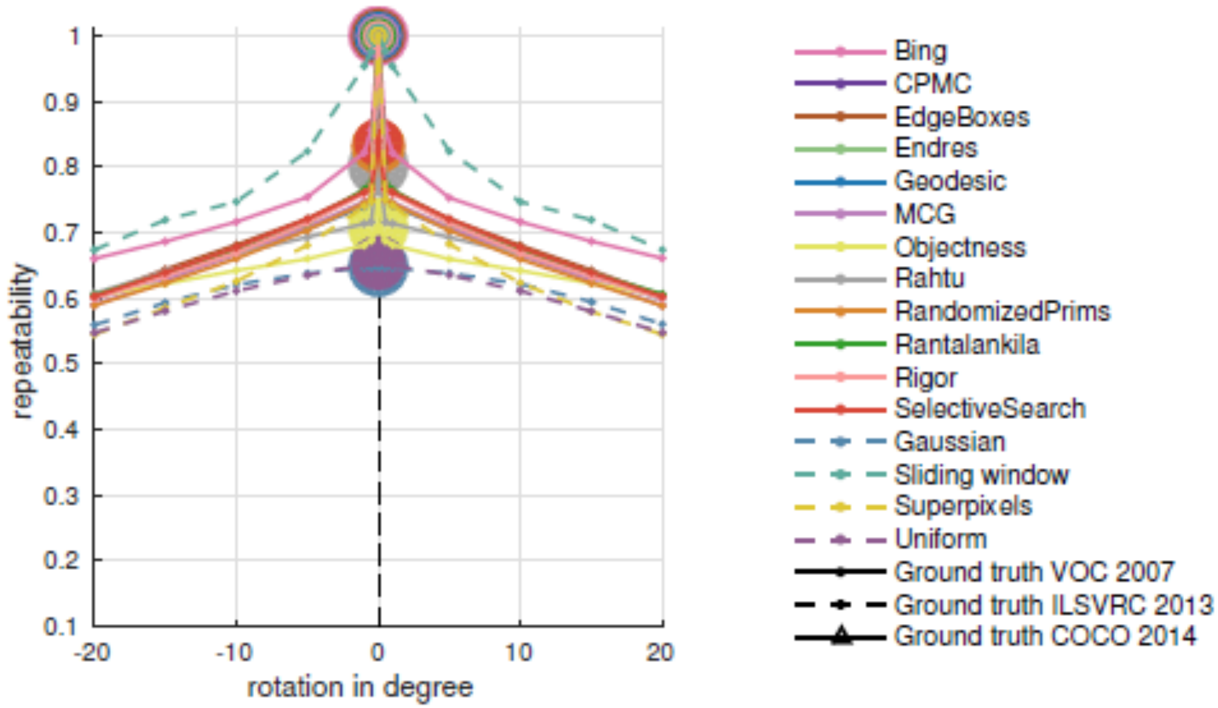


JPEG artefacts.

- **<u>Rotation:</u>**

All proposal methods are affected by image rotation.

The repeatability loss is due to matching rotated bounding boxes.

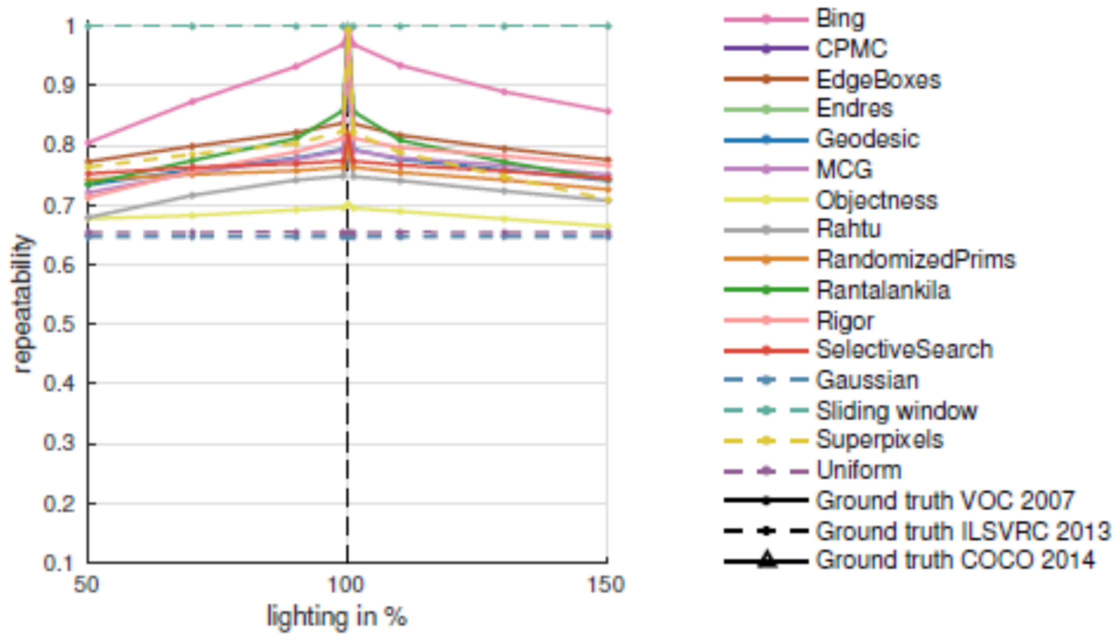


(e) Rotation.

• **Illumination:**

Methods based on superpixels are heavily affected.

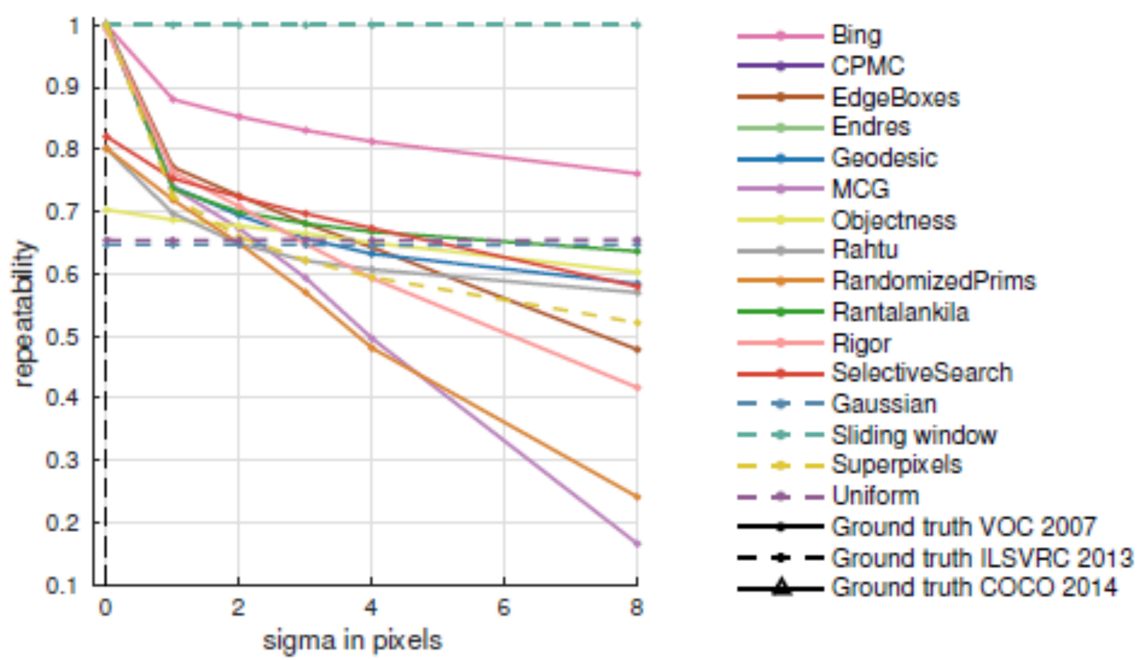
Bing is more robust, likely due to use of gradient information which is known to be fairly robust to illumination changes.



(f) Illumination.

- **Blur:**

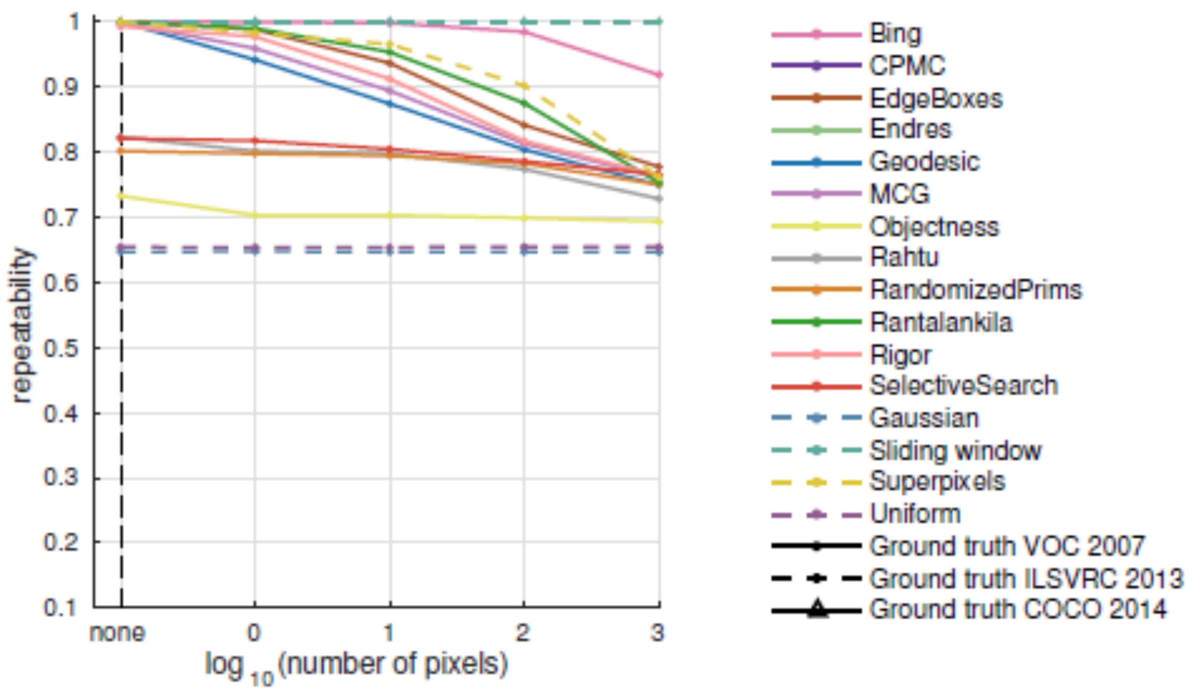
The repeatability results again exhibit a similar trend although the drop is stronger (in comparison with other effects) for a small.



(g) Blur.

- **Salt and pepper noise:**

Significant degradation in repeatability for the majority of the methods occurs when merely ten pixels are modified.



(h) Salt and pepper noise.

4. Proposal Recall

- If repeatability is a concern, the proposal method should be selected with care.

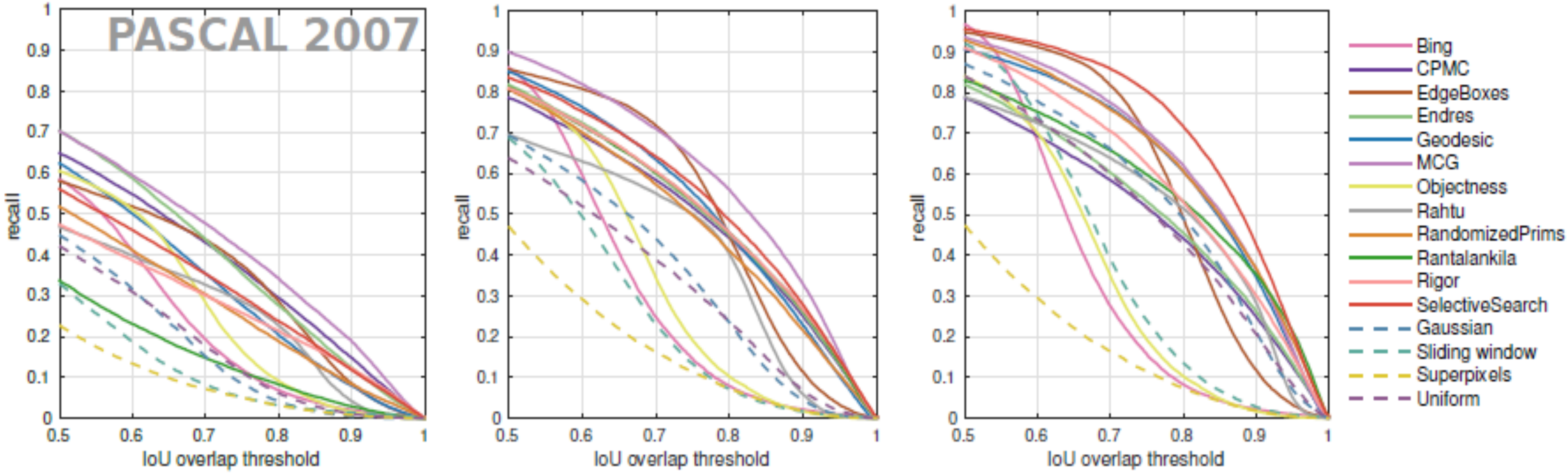
- For object detection, another aspect of interest is recall.

Dataset	Description
1. PASCAL	It includes 20 object categories that are presented in nearly 5000 unconstrained images.
2. ImageNet	In larger ImageNet 2013, there are 200 categories in over 20,000 images. Different types of objects are included that are not in PASCAL. ImageNet and PASCAL have the same number of objects/image and size of objects.
3. MS COCO	Microsoft Common objects in Context (MS COCO) has more objects/image, smaller objects, but fewer object classes (80 object categories).

Overall, the methods fall into two groups:

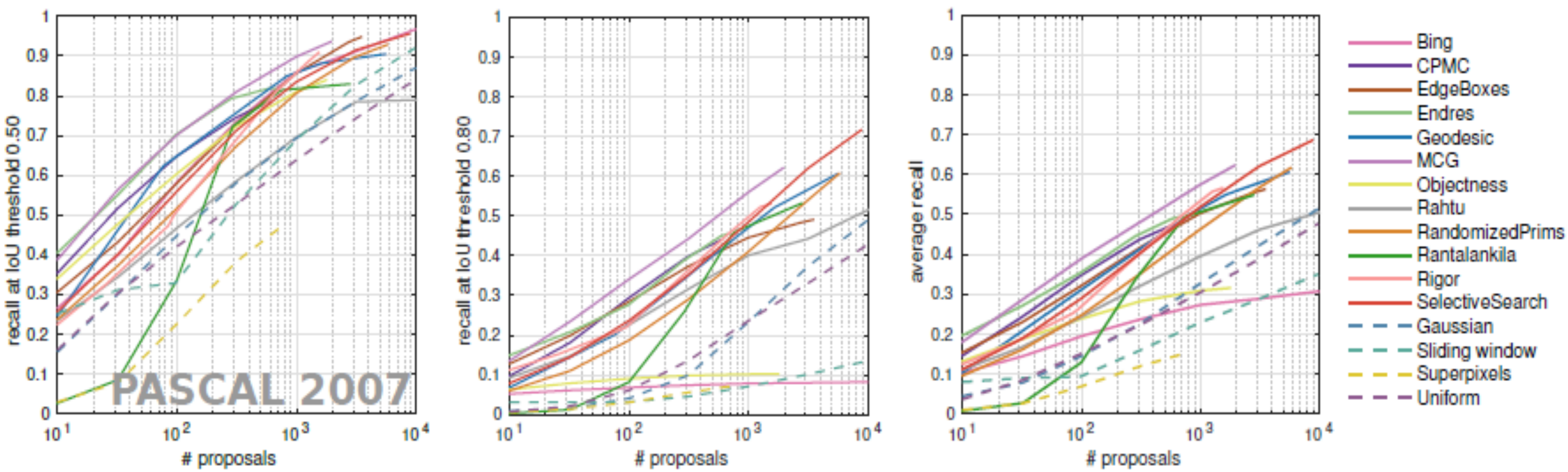
1. Well localized methods that gradually lose recall as the IoU threshold increases.

2. Methods that only provide coarse bounding box locations, so their recall drops rapidly.



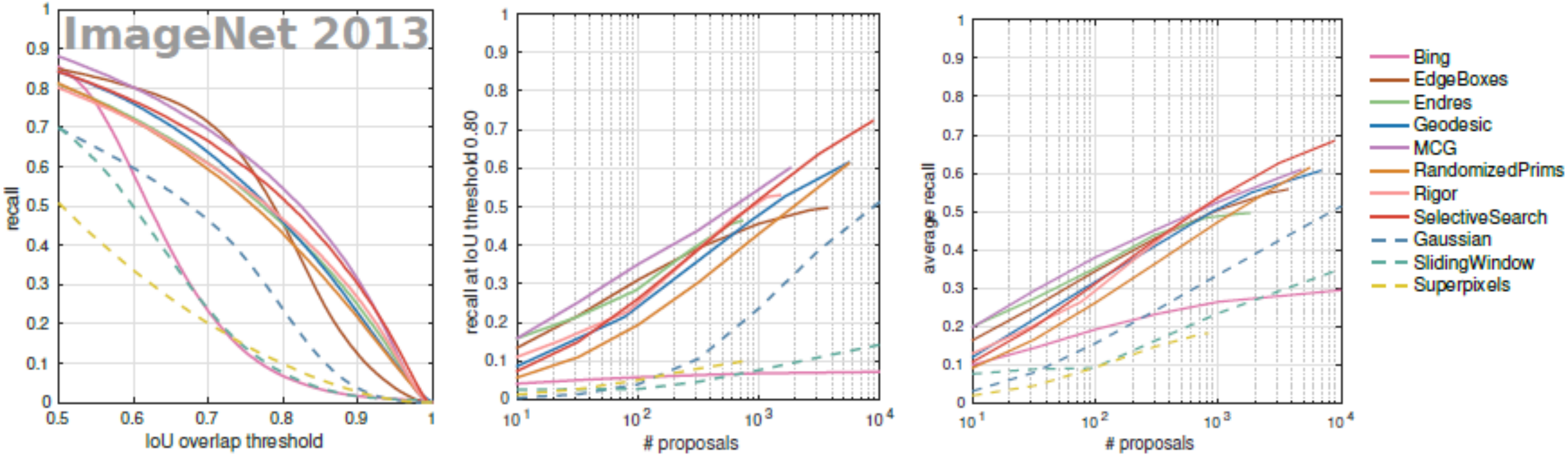
(a) 100 proposals per image. (b) 1 000 proposals per image. (c) 10 000 proposals per image.

Figure 6: Recall versus IoU threshold on the PASCAL VOC 2007 test set.



(a) Recall at 0.5 IoU. (b) Recall at 0.8 IoU. (c) Average recall (between [0.5, 1] IoU).

Recall versus number of proposal windows on the PASCAL VOC 2007 test set.

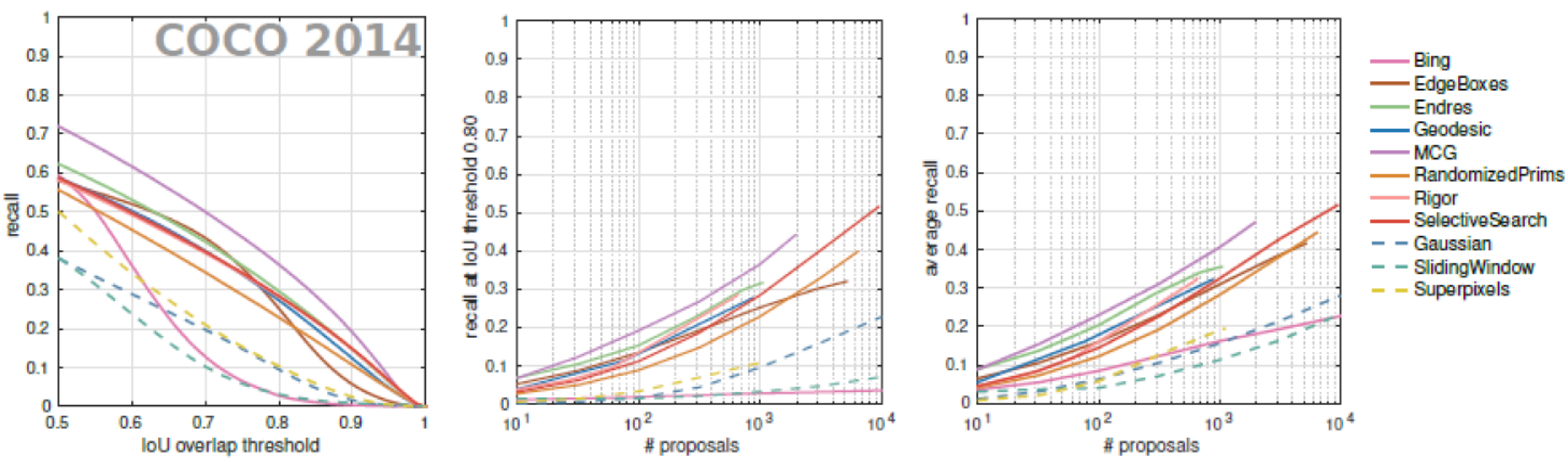


(a) 1 000 proposals per image.

(b) Recall at 0.8 IoU.

(c) Average recall (between [0.5, 1] IoU).

Recall on the ImageNet 2013 validation set.

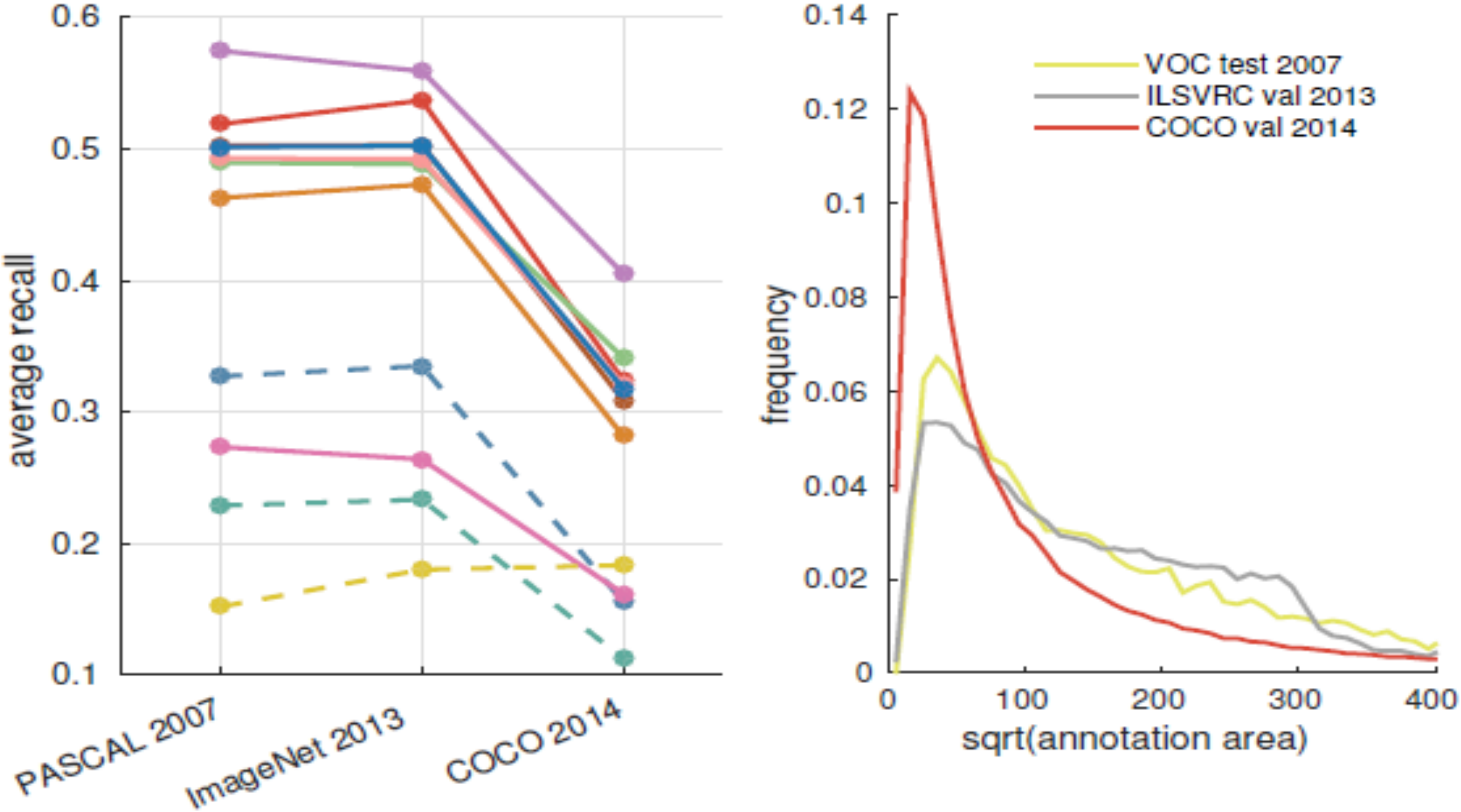


(a) 1 000 proposals per image.

(b) Recall at 0.8 IoU.

(c) Average recall (between [0.5, 1] IoU).

Recall on the MS COCO 2014 validation set.



(a) AR with 1000 proposals (b) Ground truth size

Comparison between all considered datasets: PASCAL VOC 2007 test set, ImageNet 2013 validation set, MS COCO 2014 validation set

5. Using The Detection Proposals

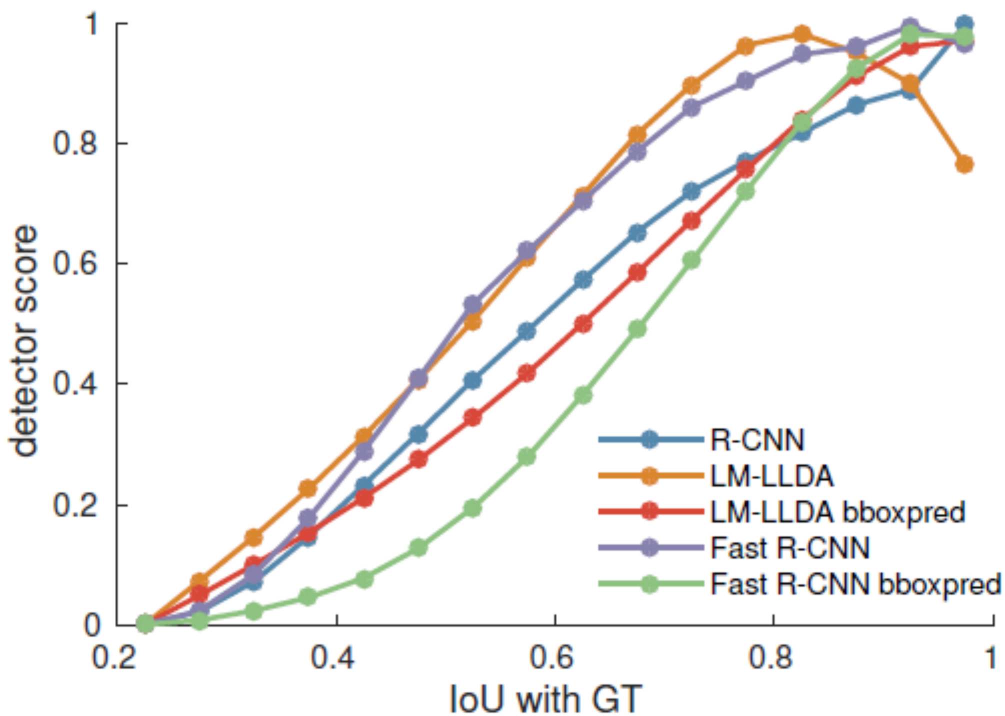
- This is an analysis of detection proposals to be used with object detection.

- The main 2 goals:

1. Measuring the performance of proposal methods for object detection.
2. The effect of object proposals metric on final detection performance.

5.1 Detector Responses Around Objects

- It is necessary to check the importance and relationship between well localized proposals (high IOU) and object detection (recall).



Normalised detector scores as a function of the overlap between the detector window and the ground truth.

5.2 LM-LLDA, R-CNN and Fast R-CNN detection performance

1. Apply LM-LLDA models to generate dense detections using the standard sliding window.
2. Apply different object proposals to filter these detections at test time.

* These steps are used to evaluate the effect of proposals on detection quality.

- Using only 1000 proposals, the detection quality is reduced.
- But, methods with high average recall (AR) also have high mean average precision (mAP), and vice versa.

Proposals	LM-LLDA
Dense	33.5/34.4
Bing	21.8/22.4
CPMC	30.0/30.7
EdgeBoxes	31.8/32.2
Endres	31.2/31.7
Geodesic	31.8/32.2
MCG	32.5/33.0
Objectness	25.0/25.4
Rahtu	29.6/30.4
RandomizedPrims	30.5/30.9
Rantalankila	30.9/31.4
Rigor	31.5/32.1
SelectiveSearch	31.7/32.3
Gaussian	27.3/28.0
Sliding window	20.7/21.5
Superpixels	11.2/11.3
Uniform	26.0/26.6

- From table below:

(1) clearly hurt performance (bicycle, boat, bottle, car, chair, horse, mbike, person), reducing the recall and precision because of bad localization.

(2) improve performance (cat, table, dog),

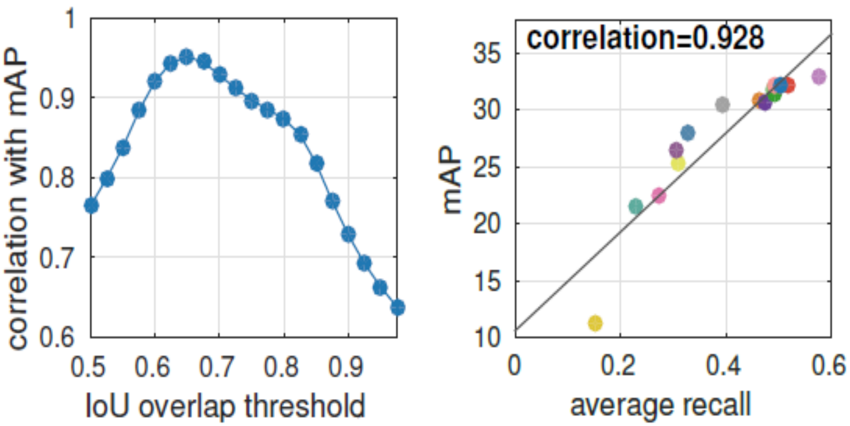
(3) do not show significant change (all remaining classes).

	aero	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mean
LM-LLDA Dense	33.7	61.3	12.4	18.5	26.7	53.0	57.2	22.4	22.7	25.6	25.1	14.0	59.2	51.0	39.1	13.6	21.7	38.0	48.8	44.0	34.4
Bing	-7.5	-23.2	-6.2	-8.1	-10.6	-13.3	-17.5	-6.8	-9.8	-15.4	-7.5	-1.4	-19.6	-19.0	-16.1	-3.4	-6.6	-18.1	-18.8	-10.0	-11.9
CPMC	-1.0	-15.0	-0.2	-4.4	-13.5	-1.8	-9.2	3.2	-9.1	-2.6	5.1	2.2	-4.2	-4.8	-7.0	-2.0	-2.6	1.2	-4.1	-4.9	-3.7
EdgeBoxes	-2.0	-6.1	-0.7	-3.8	**-6.7**	**0.6**	**-5.8**	-1.1	**-2.0**	-1.8	-4.6	0.4	-1.3	**-1.3**	-3.0	-1.7	**-0.1**	-0.9	-0.2	-1.1	**-2.2**
Endres	-1.5	**-5.8**	-0.6	-4.8	-12.7	-1.1	-7.1	3.4	-6.9	-3.2	4.7	1.9	-2.4	-2.4	-7.7	-2.8	-1.9	1.5	0.4	-4.2	-2.7
Geodesic	-1.9	-8.1	-0.2	-4.6	-14.4	0.6	-6.5	2.6	-7.3	**-1.3**	4.7	2.4	-2.5	-2.7	-4.7	-1.2	-0.7	-0.1	**1.9**	0.2	**-2.2**
MCG	-0.7	-7.2	0.1	-3.6	**-6.7**	-1.2	-7.0	3.4	-3.2	-2.3	5.0	1.9	-3.5	**-1.3**	**-1.5**	**-1.1**	-1.3	2.2	0.3	**0.5**	**-1.4**
Objectness	-10.3	-15.1	-2.0	-6.2	-11.0	-9.5	-13.0	-3.6	-10.0	-6.4	-7.8	-1.0	-11.6	-15.9	-13.0	-2.7	-5.8	-11.2	-10.9	-12.9	-9.0
Rahtu	-0.3	-13.2	-0.3	**-1.2**	-13.0	-0.6	-12.0	3.3	-10.5	-4.3	2.0	2.1	-3.2	-4.9	-7.9	-2.8	-4.9	-5.0	0.0	-3.7	-4.0
Rand.Prim	**2.1**	-10.4	-0.5	-4.5	-13.2	-1.9	-10.1	5.0	-6.7	-3.5	2.0	2.4	-4.4	-5.1	-10.0	-2.3	-1.8	1.2	-3.8	-4.4	-3.5
Rantalankila	0.5	-13.6	0.3	-3.0	-12.9	-3.6	-9.0	4.4	-5.6	-3.7	4.1	2.5	-2.2	-4.0	-7.8	-2.5	-3.8	**2.1**	-1.5	-0.7	-3.0
Rigor	1.7	-7.9	0.5	-4.1	-12.4	-0.8	-9.0	**6.3**	-6.9	-1.7	1.8	**2.9**	**-0.9**	-3.3	-7.7	-1.8	-1.3	1.6	-1.2	-1.7	**-2.3**
SelectiveSearch	1.3	-7.7	**1.0**	-4.3	-11.1	-1.7	-7.8	3.9	-4.8	-1.5	**5.4**	2.2	-1.4	-3.8	-6.0	-1.5	-0.8	0.6	-2.4	-2.1	**-2.1**
Gaussian	-6.6	-13.4	-0.7	-4.4	-15.0	-6.1	-16.0	0.9	-9.1	-8.0	0.3	1.2	-4.2	-6.9	-10.3	-2.3	-6.5	-4.5	-3.6	-12.1	-6.4
SlidingWindow	-21.8	-20.7	-3.2	-8.1	-16.6	-14.7	-22.1	-0.7	-9.8	-11.7	-10.2	-1.4	-14.7	-20.1	-14.8	-3.8	-7.7	-21.0	-20.8	-14.8	-12.9
Superpixels	-23.9	-52.2	-3.1	-9.4	-17.4	-43.9	-42.3	-10.2	-11.3	-12.6	-15.8	-8.5	-50.1	-41.7	-30.9	-4.4	-10.6	-25.2	-39.7	-8.2	-23.1
Uniform	-3.2	-18.8	-4.0	-4.8	-15.2	-8.6	-16.6	0.2	-10.4	-8.8	3.7	1.3	-6.6	-11.3	-10.2	-3.6	-8.9	-5.8	-5.1	-20.2	-7.8
Top methods avg.	-0.3	-7.4	0.1	-4.1	-10.2	-0.5	-7.2	3.0	-4.8	-1.7	2.5	2.0	-1.9	-2.5	-4.6	-1.5	-0.8	0.7	-0.3	-0.8	-2.0

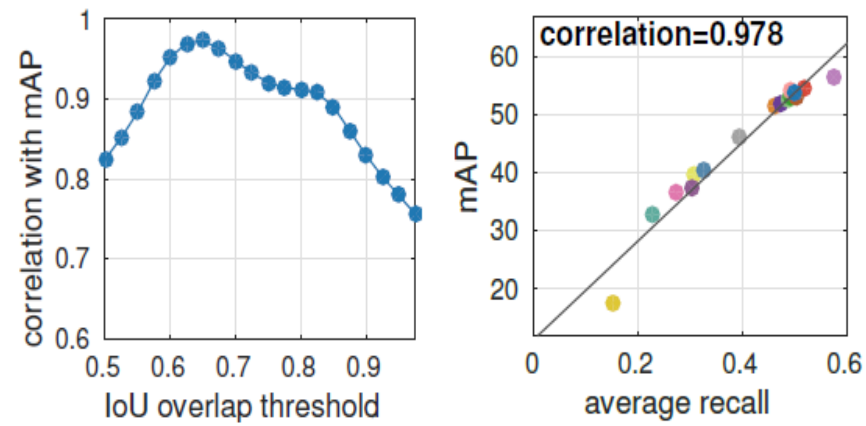
- Fast R-CNN after re-training for each method.
- In the right most column, Fast R-CNN trained with 1000 SelectiveSearch proposals and applied at test time with a given proposal method, versus Fast R-CNN trained for the test time proposal method.

Proposals	LM-LLDA	R-CNN	Fast R-CNN	ΔTrain
Dense	33.5/34.4	–	–	–
Bing	21.8/22.4	36.7	37.3/49.0	+6.3
CPMC	30.0/30.7	51.7	53.7/57.1	-1.3
EdgeBoxes	31.8/32.2	53.0	55.4/**60.4**	+3.3
Endres	31.2/31.7	52.8	54.2/57.4	-0.2
Geodesic	31.8/32.2	53.8	53.6/57.5	-0.4
MCG	**32.5/33.0**	**56.5**	**58.1**/60.3	+1.8
Objectness	25.0/25.4	39.7	41.5/51.4	+9.1
Rahtu	29.6/30.4	46.1	48.9/53.6	+0.7
RandomizedPrims	30.5/30.9	51.6	53.2/57.6	-0.6
Rantalankila	30.9/31.4	53.1	55.0/57.9	-0.5
Rigor	31.5/32.1	54.1	55.4/58.4	-0.2
SelectiveSearch	31.7/32.3	54.6	56.3/59.5	+0.0
Gaussian	*27.3/28.0*	*40.6*	*44.6/50.8*	+0.8
Sliding window	20.7/21.5	32.7	32.7/44.8	+3.3
Superpixels	11.2/11.3	17.6	15.4/20.3	-2.0
Uniform	26.0/26.6	37.3	39.5/46.9	-0.1

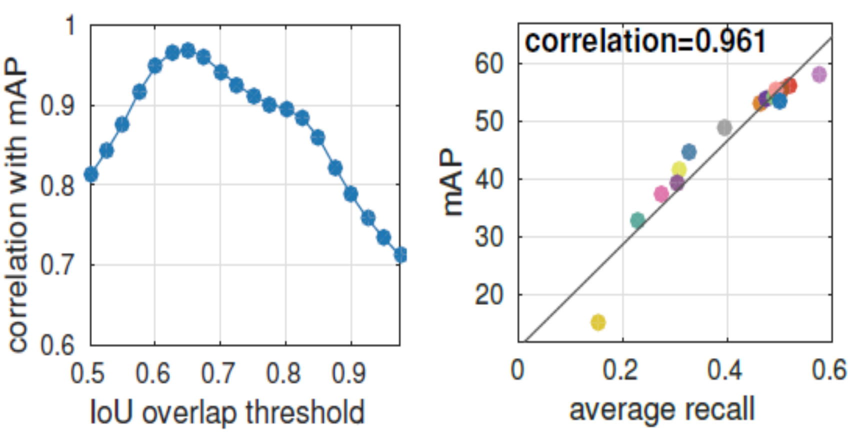
5.3 Predicting detection performance



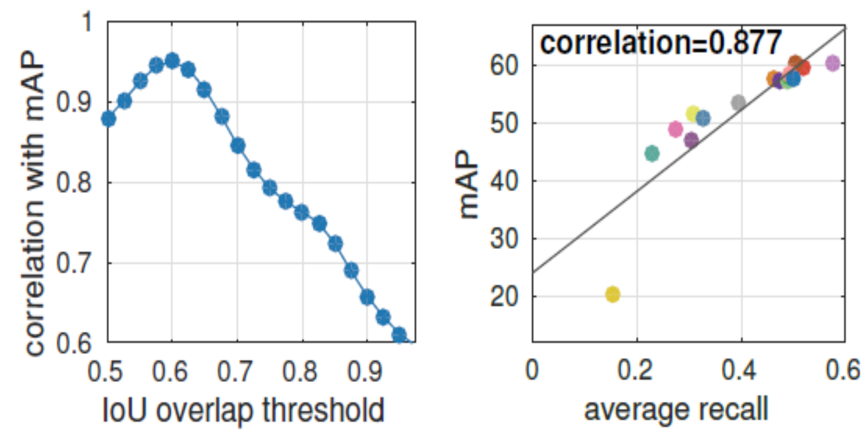
(a) LM-LLDA with bounding box regression



(b) R-CNN without bounding box regression



(c) Fast R-CNN without bounding box regression



(d) Fast R-CNN with bounding box regression

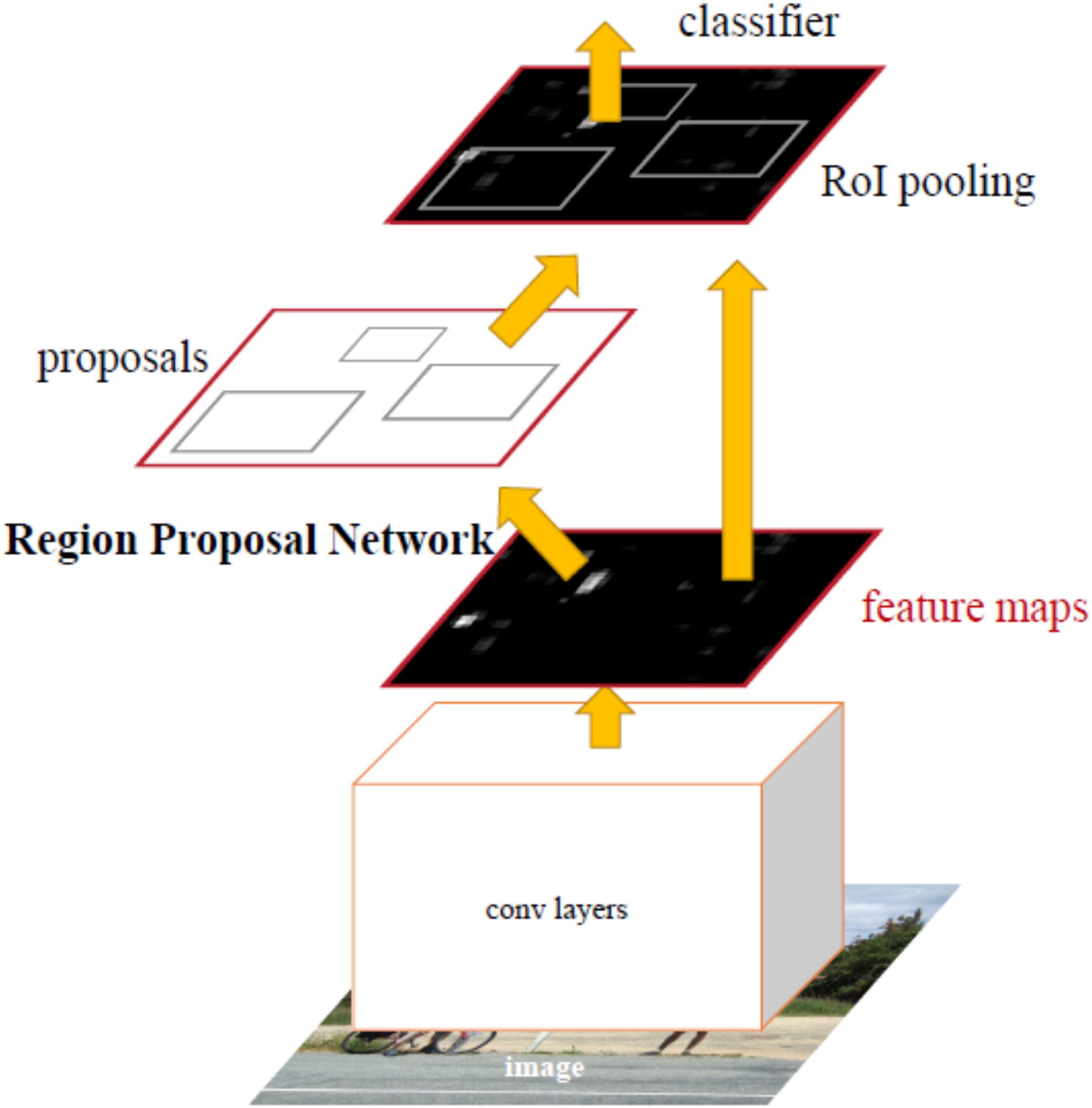
Correlation between detector performance on PASCAL 07 and different proposal metrics. Left columns: correlation between mAP and recall at different IoU thresholds. Right columns: correlation between mAP and AR.

Related Work:

Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

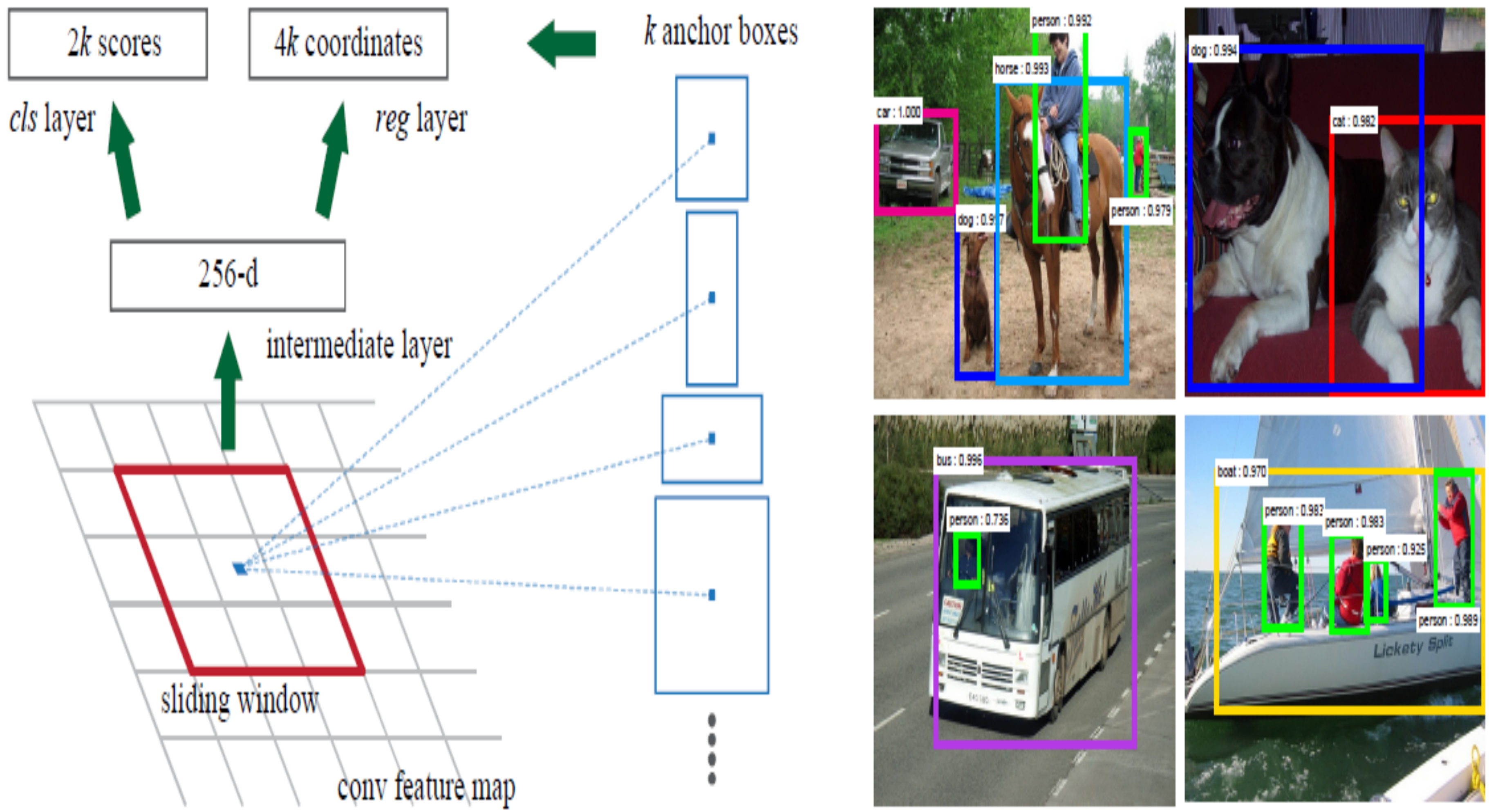
Shaoqing Ren[1], Kaiming He[2], Ross Girshick, and Jian Sun[3]
[1]University of Science and Technology of China
[2]Microsoft Research
[3]Facebook AI Research

- This object detection system is composed of two modules. The first module is a deep fully convolutional network that proposes regions, and the second module is the Fast R-CNN detector that uses the proposed regions.
- The RPN module tells the Fast R-CNN module where to look.
- A Region Proposal Network (RPN) takes an image (of any size) as input and outputs a set of rectangular object proposals, each with an objectness score.



Faster R-CNN is a single, unified network for object detection. The RPN module serves as the 'attention' of this unified network.

- For region proposals generation, slide a small network over the convolutional feature map output by the last shared convolutional layer.
- This small network takes as input an nxn spatial window of the input convolutional feature map.
- Each sliding window is mapped to a lower-dimensional feature (256-d for ZF and 512-d for VGG, with ReLU following).
- This feature is fed into two sibling fully connected layers—a box-regression layer (reg) and a box-classification layer (cls).



Left: Region Proposal Network (RPN). **Right**: Example detections using RPN proposals on PASCAL VOC 2007 test. Our method detects objects in a wide range of scales and aspect ratios.

Conclusion

- This paper revisits the majority of existing detection proposal methods, proposed new evaluation metrics, and performed an extensive and direct comparison of existing methods.
- The repeatability of all proposal methods is limited: small changes to an image cause a noticeable change in the set of produced proposals.
- For object detection, improving proposal localization accuracy (improved IoU) is as important as improving recall.
- To simultaneously measure both proposal recall and localization accuracy, average recall (AR) summarizes the distribution of recall across a range of overlap thresholds.

Strengths

- This paper provides a new metric, Average Recall (AC), that relates between accuracy (recall) and good localization (IOU).
- It demonstrates different evaluation protocol to compare between proposal methods (repeatability, recall and using proposal methods for object detection).

Weaknesses

- This paper depends only on 12 proposal methods, because their implementations are available.
- The baseline proposal methods are not algorithms (uniform, Gaussian, sliding window and superpixels).

Overall Rating

- **My Rating Scale (0-5): 1**

The new performance metric which is Average Recall (AC) is just an Average Best Overlap (ABO) within range 0.5:1

Comparison is taken place between 12 proposal methods only.