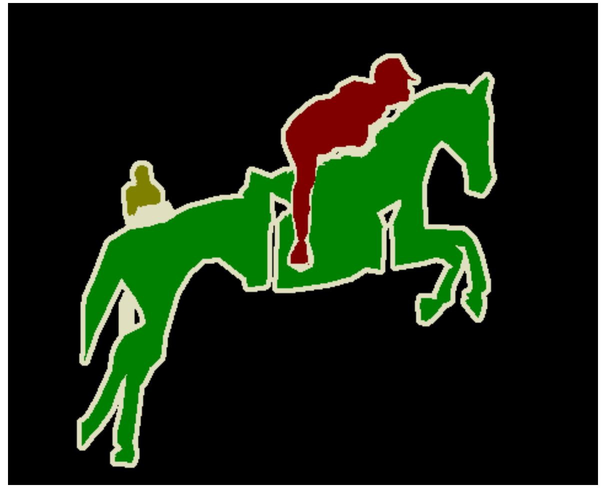
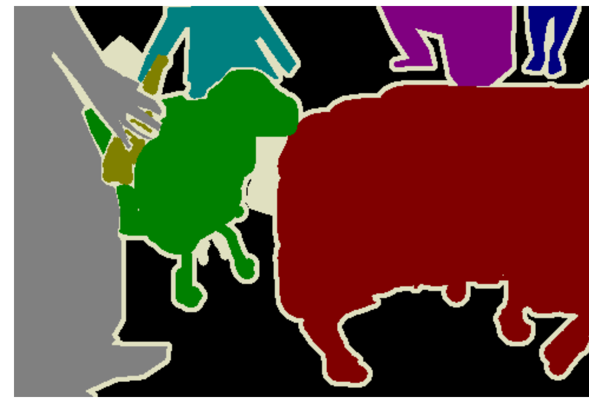




Object Proposal

Siyang Li

Advisor:C-C Jay Kuo

Outline

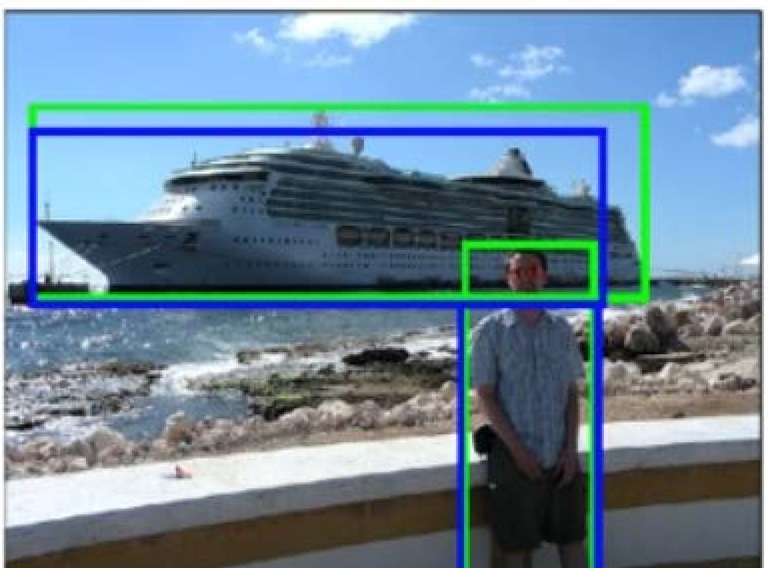
- Problem definition
- Motivation
- History
- Challenges
- Evaluation
- Popular approaches
- Recent work
- Future research

Problem definition

- What is object proposal?
  - Find a small set of **regions** that are highly likely to contain objects
  - The proposed regions are **category-independent usually, but category-dependent proposals are possible**

Problem definition

- Proposals as bounding boxes
- Proposals as segmentation masks

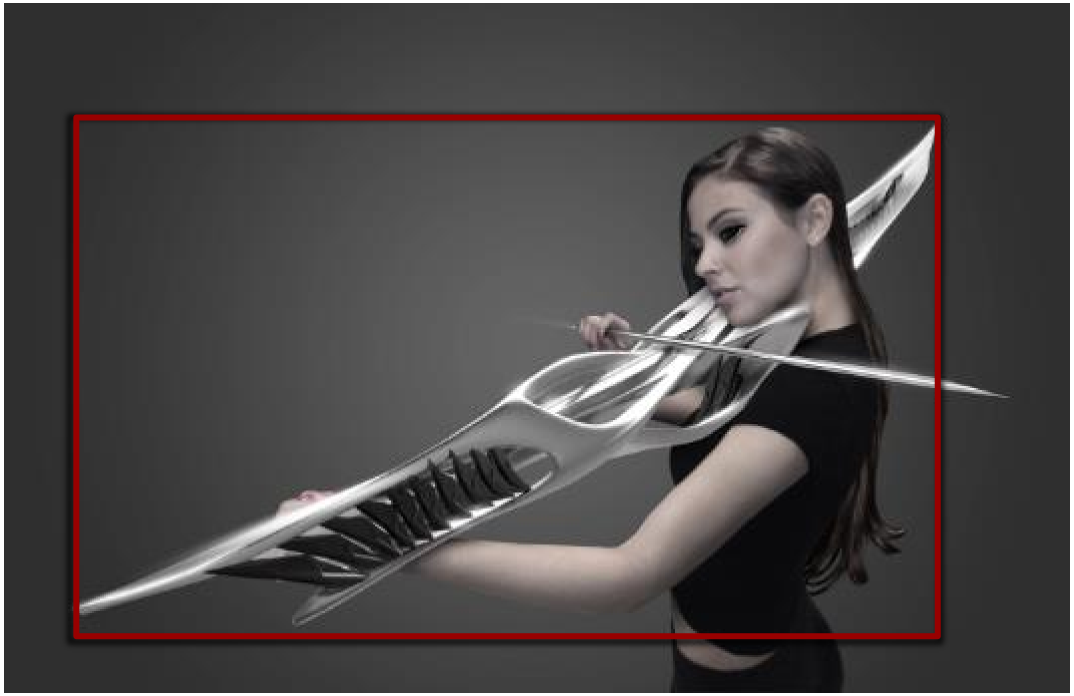


Motivation

- Observations
  - All object of interest **share common visual properties** that distinguish them from the background
  - Human have the ability to localize objects without recognizing them

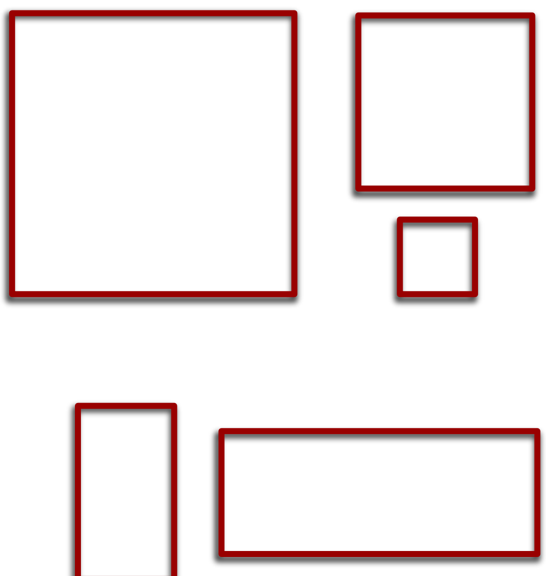
Motivation

- In the following images, you can find the strange instrument as an object without knowing what it exactly is.



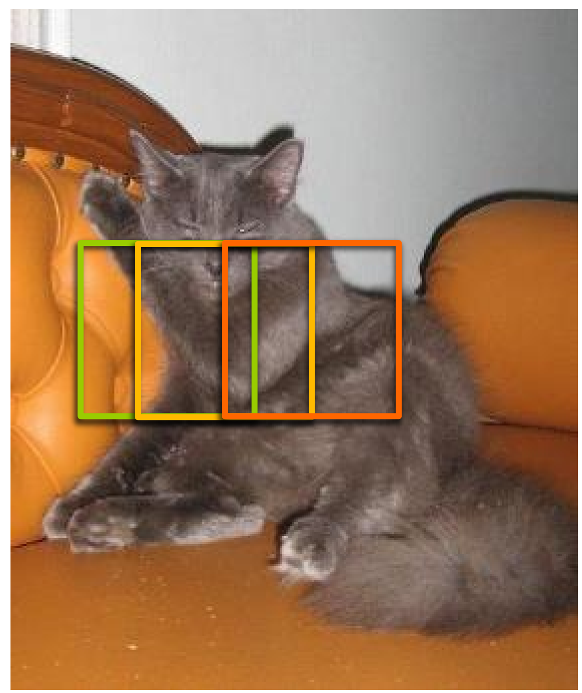
Motivation

- Problems with traditional detectors
  - Sliding windows paradigm
  - About a million windows to evaluate per image



scale

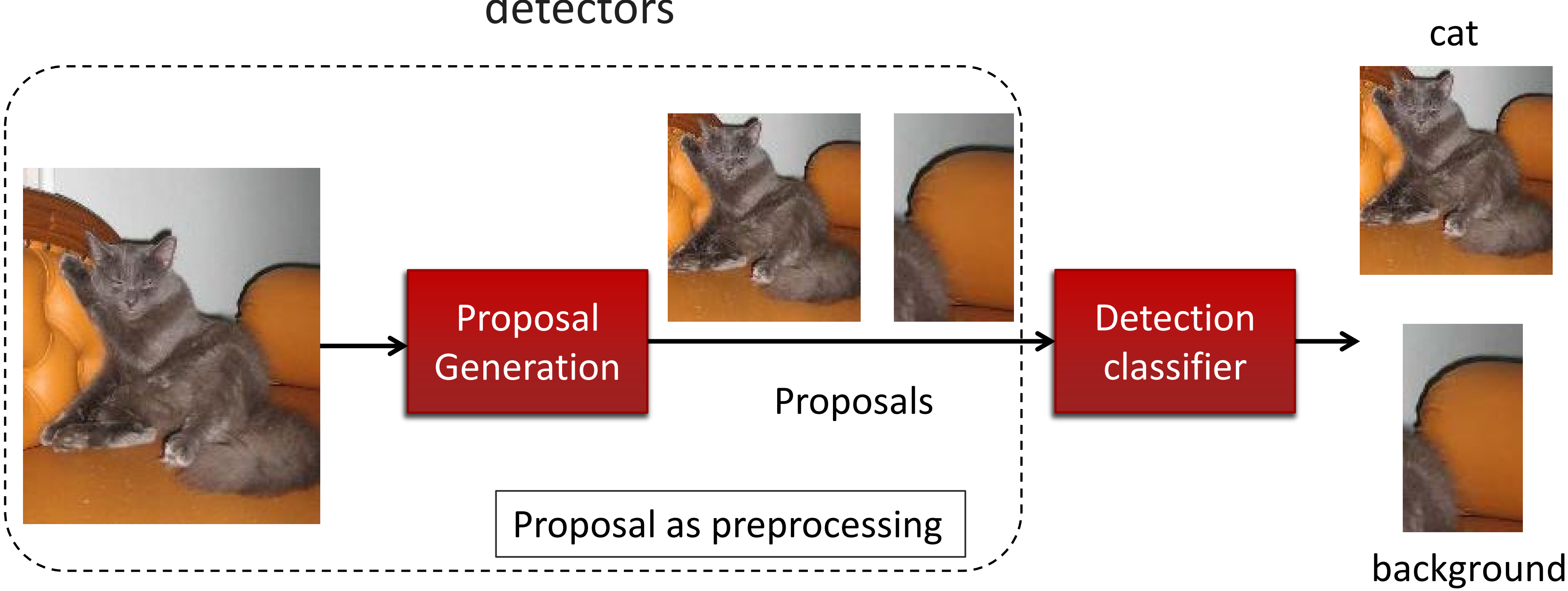
aspect ratio



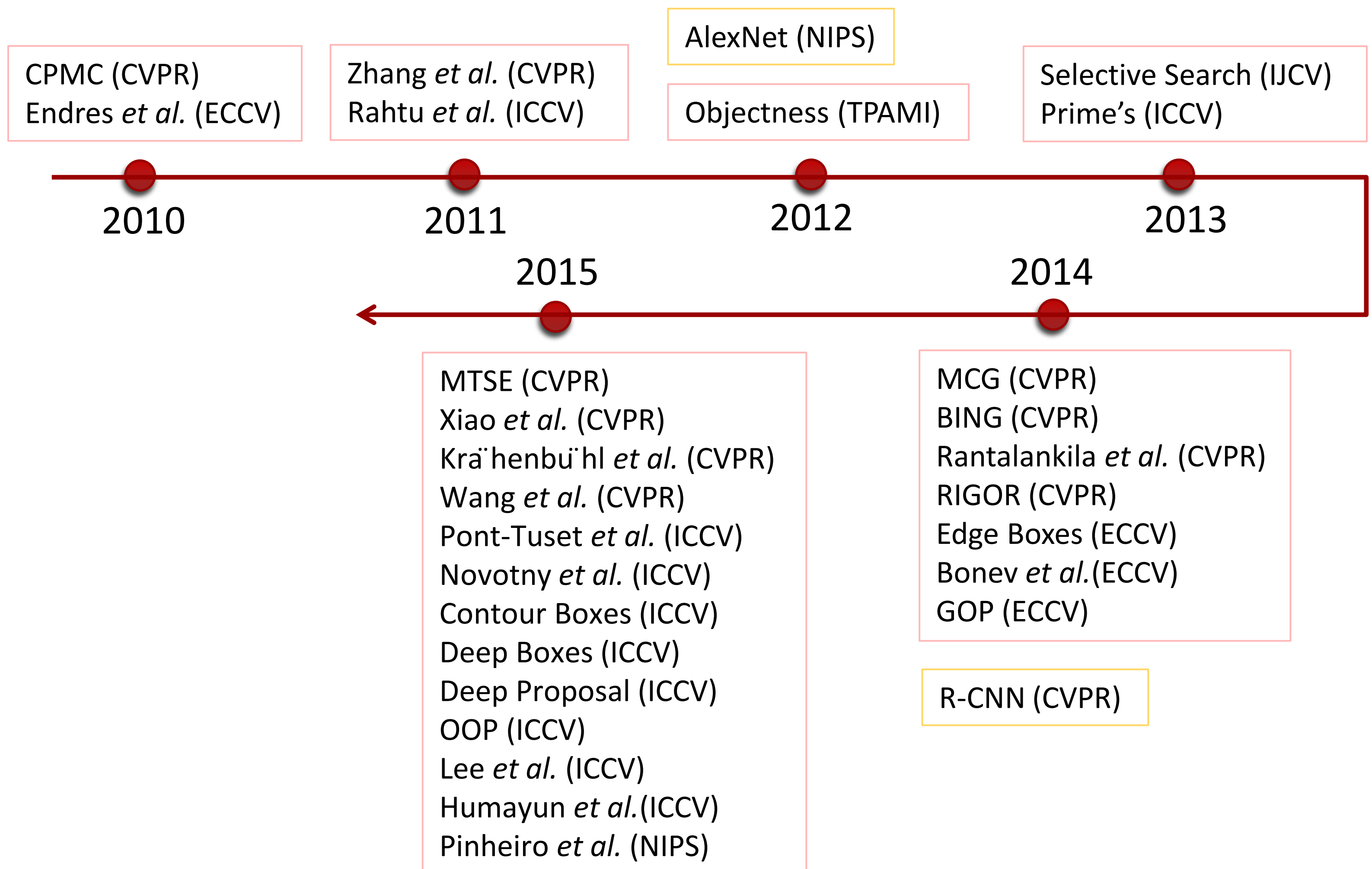
location

Motivation

- Object proposal serves as a pre-processing step
  - Do not propose obvious false window
  - Enable more complex classifier such as CNN detectors



History


CPMC (CVPR)
Endres et al. (ECCV)
2010
Zhang et al. (CVPR)
Rahtu et al. (ICCV)
2011
AlexNet (NIPS)
Objectness (TPAMI)
2012
Selective Search (IJCV)
Prime's (ICCV)
2013
2014
MCG (CVPR)
BING (CVPR)
Rantalankila et al. (CVPR)
RIGOR (CVPR)
Edge Boxes (ECCV)
Bonev et al.(ECCV)
GOP (ECCV)
R-CNN (CVPR)
2015
MTSE (CVPR)
Xiao et al. (CVPR)
Krähenbühl et al. (CVPR)
Wang et al. (CVPR)
Pont-Tuset et al. (ICCV)
Novotny et al. (ICCV)
Contour Boxes (ICCV)
Deep Boxes (ICCV)
Deep Proposal (ICCV)
OOP (ICCV)
Lee et al. (ICCV)
Humayun et al.(ICCV)
Pinheiro et al. (NIPS)

Challenges

- Size variation

Small objects

Large objects

Challenges

- Illumination change

Challenges

- Occlusion

Person, chair

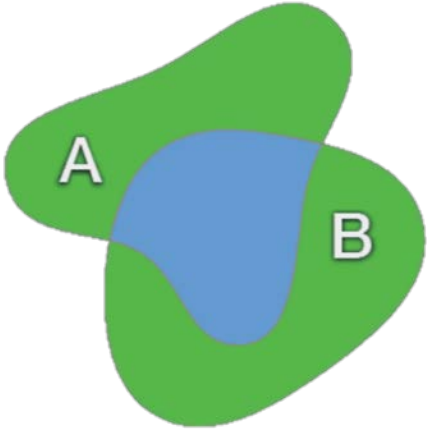
Person, laptop, bed

Evaluation

- A good proposal approach should
  - Achieve **high recall** (detection rate) given required localization accuracy
  - Generate a **small number** of proposals (~1000)
  - Obtain a **high computation efficiency**
  - Generalize **across object categories**

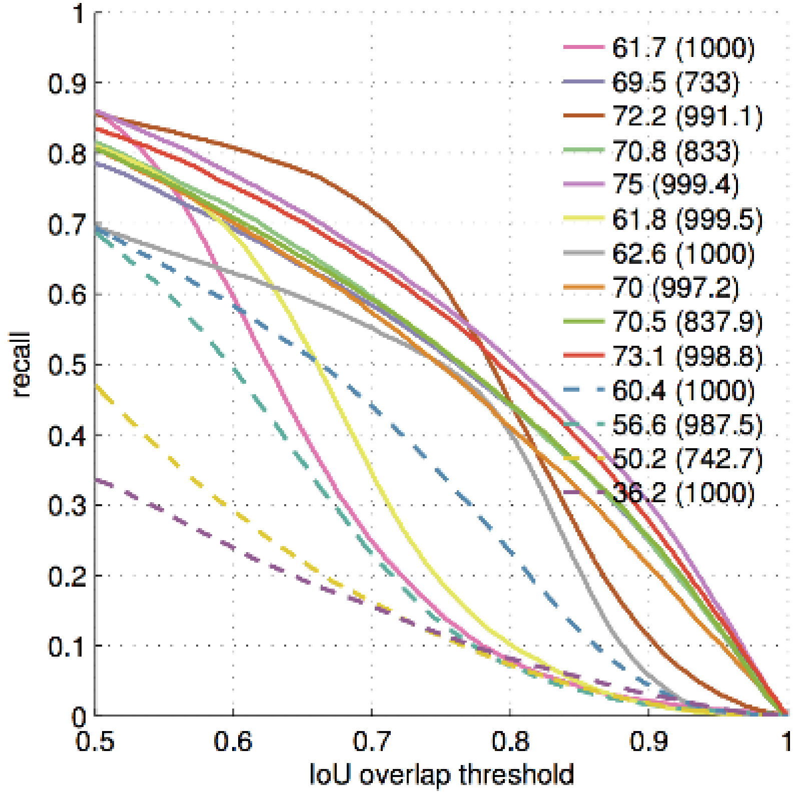
Evaluation

- Recall (detection rate)
  - Overlap (Localization accuracy) $IoU(A,B) = \frac{A \bigcap B}{A \bigcup B}$
  - Best overlap $b(O_k) = \max_P \{IoU(O_k, P)\}$
  - Average best overlap $\frac{1}{N}\sum_k b(O_k)$
  - Recall@alpha $\frac{1}{N}\sum_k [b(O_k) > \alpha]$



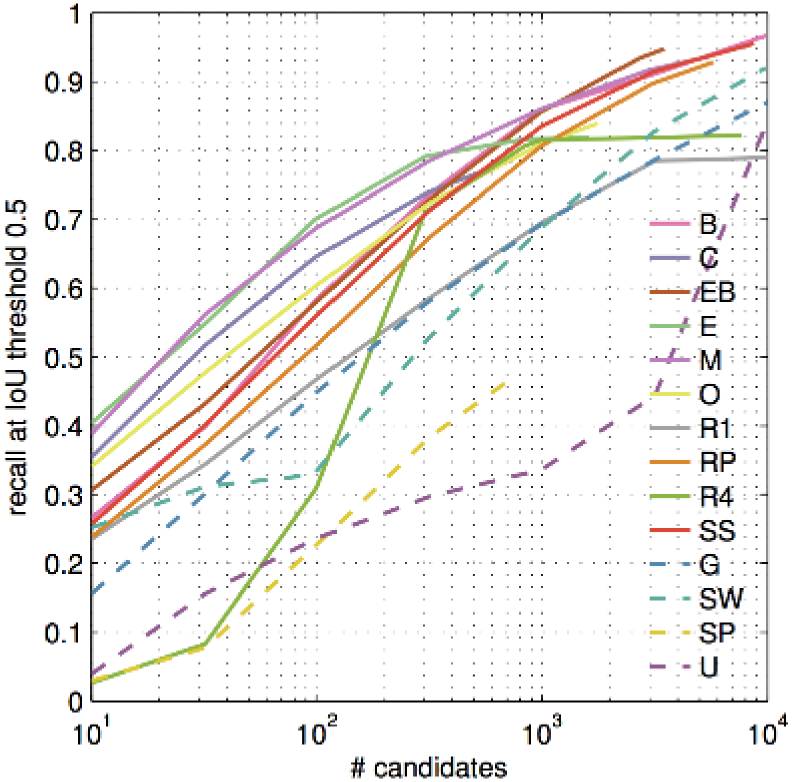
Evaluation

- Accuracy of localization
  - **Recall vs. alpha** with a fixed number of proposals



Evaluation

- The number of proposals
  - **Recall**@fixed alpha **vs. #proposals**



Evaluation

- Computation efficiency
  - # seconds needed for an image (secs/image)
  - # images can be processed in one second (FPS)

Evaluation

- Generalization across objects
  - Split the dataset: use half of categories for training, the other half for testing
  - Train on PASCAL and test on COCO

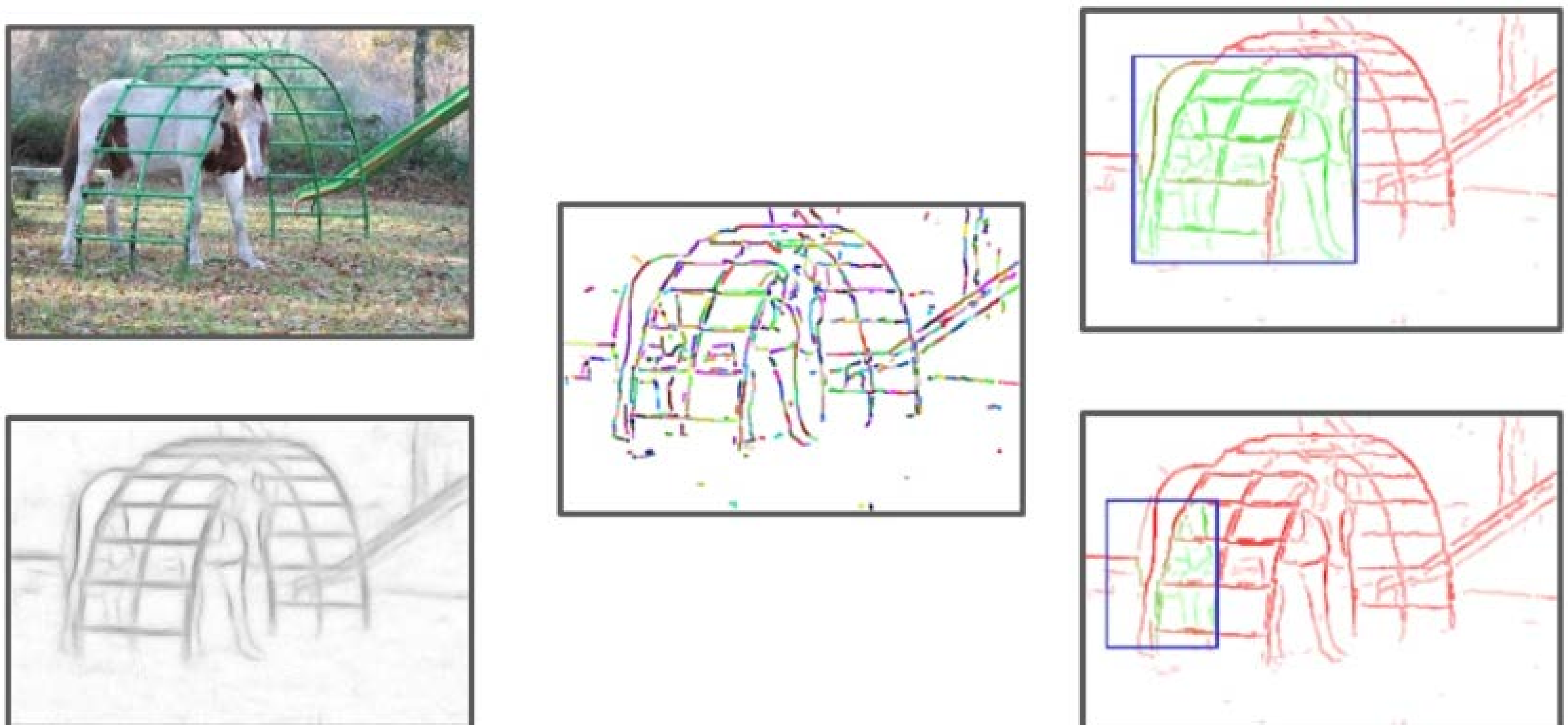
Popular approaches

- Bounding box based
  - BING (CVPR'14)
  - **Edge Boxes (ECCV'14)**
- Segmentation based
  - **Selective Search (IJCV'13)**
  - MCG (CVPR'14)

Edge Boxes

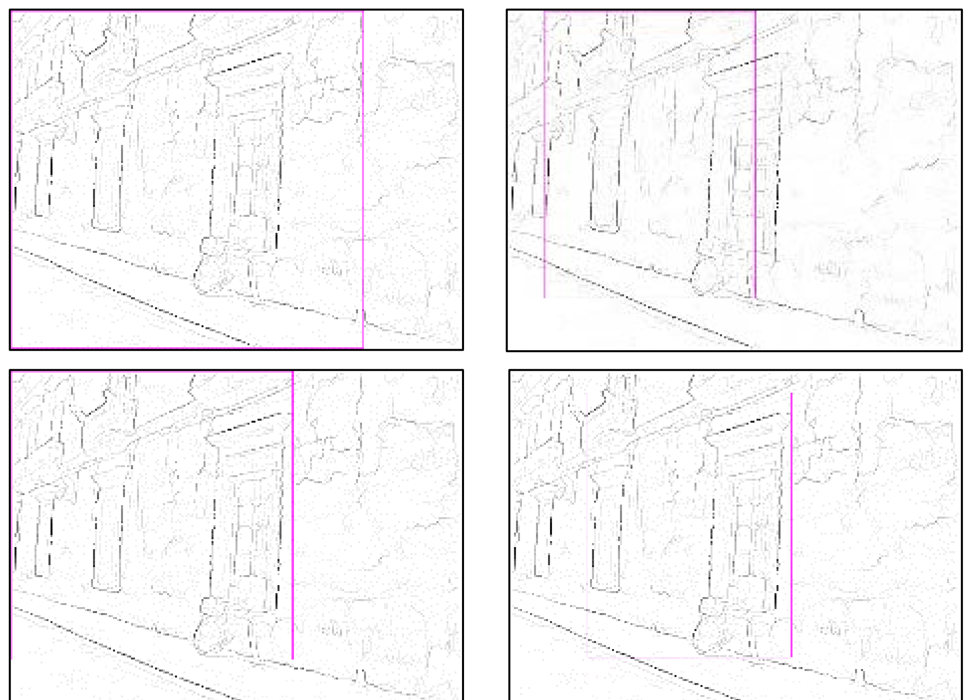
- Edge detection (Structured Edge detector)
- Edge grouping in to edge pieces
- Score a window based on the pattern of edge
  - How many pieces are in the box?
  - How many pieces are straddling the box?

Edge Boxes



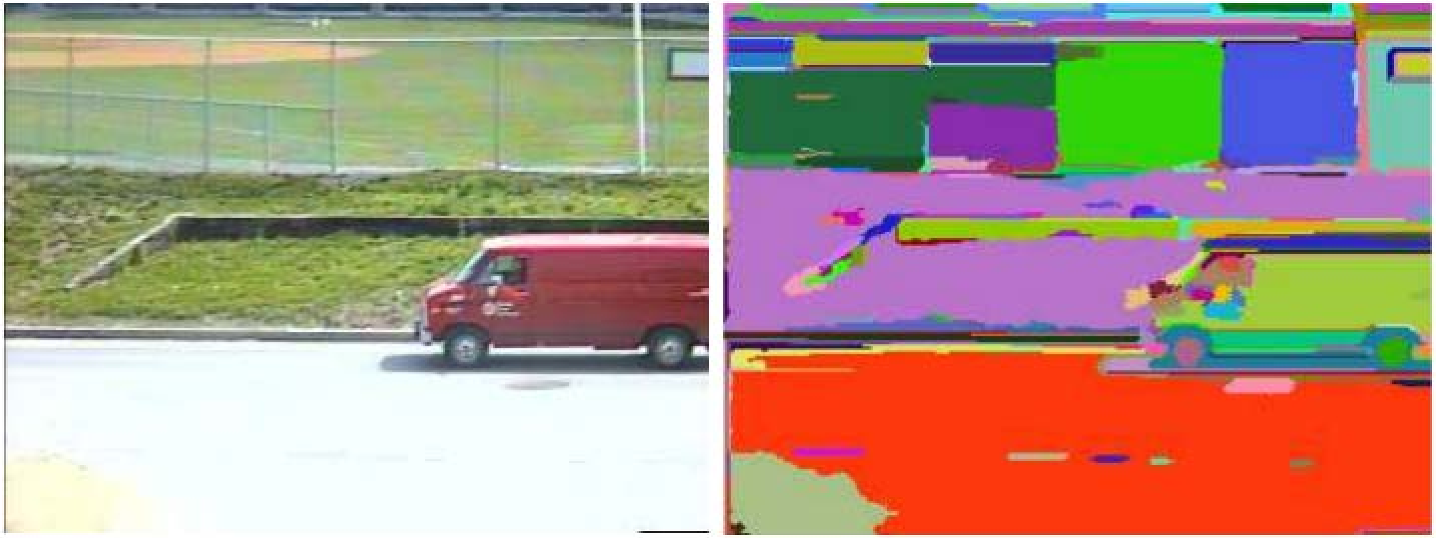
Edge Boxes

- Fast (~0.4secs/image)
- Biased towards larger boxes



Selective Search

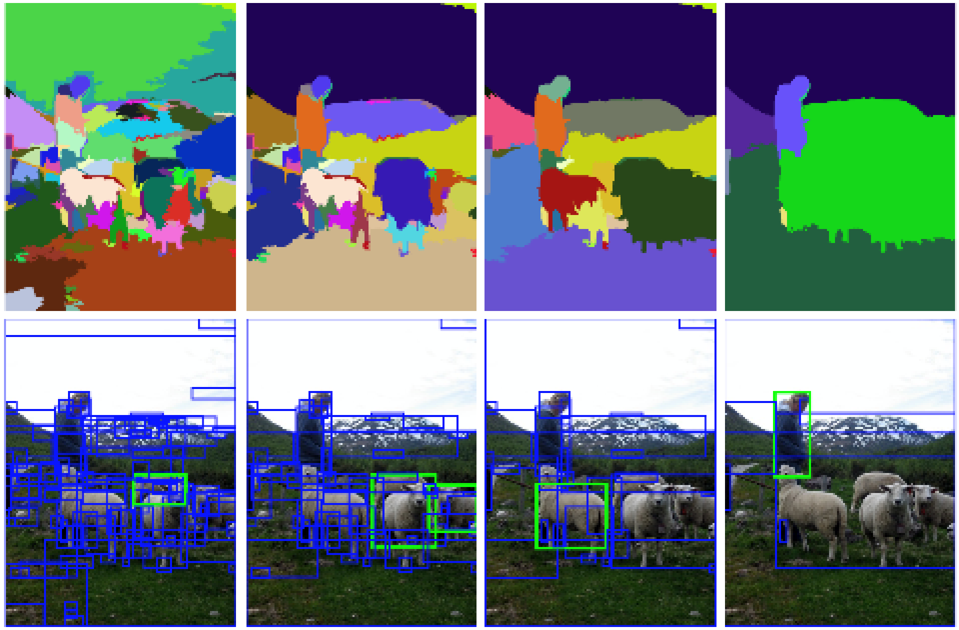
- Generate superpixels on image of multiple scales
  - FH segmentation algorithm [1]
  - Group connected pixels with similar color



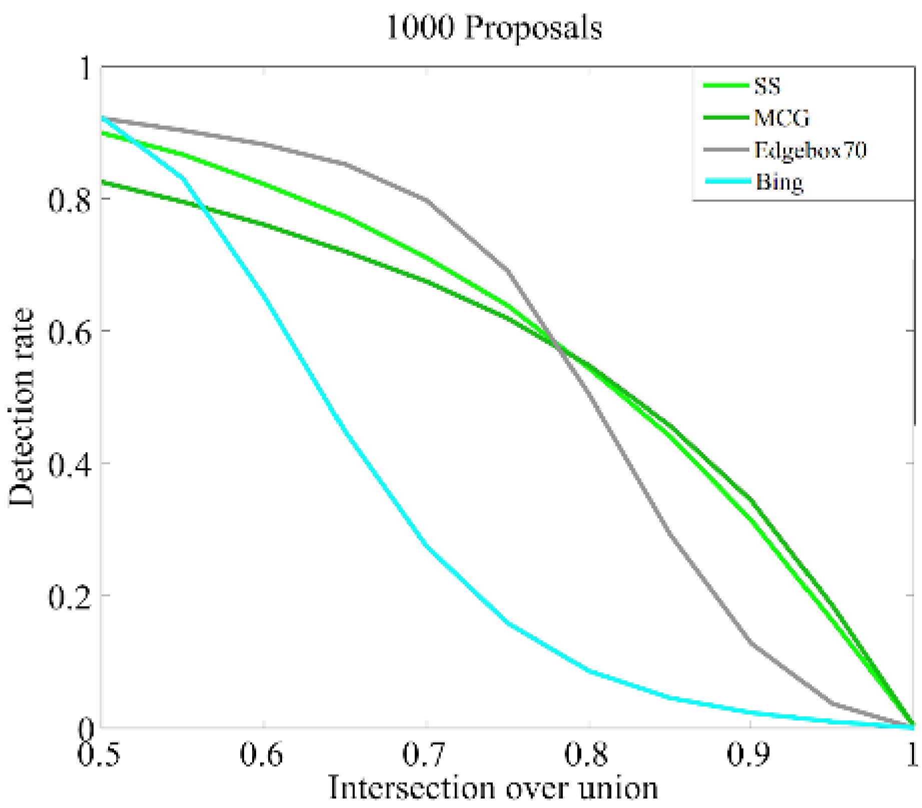
[1] Efficient Graph-Based Image Segmentation P. Felzenszwalb, D. Huttenlocher International Journal of Computer Vision, Vol. 59, No. 2, September 2004

Selective Search

- Generate superpixels on image of multiple scales
- Group similar superpixels to generate proposals
  - Similarity measured on low-level features



Summary



Approaches	Speed (secs/image)
BING	0.2
Edge Boxes	0.25
Selective Search	~10
MCG	34.3

Summary

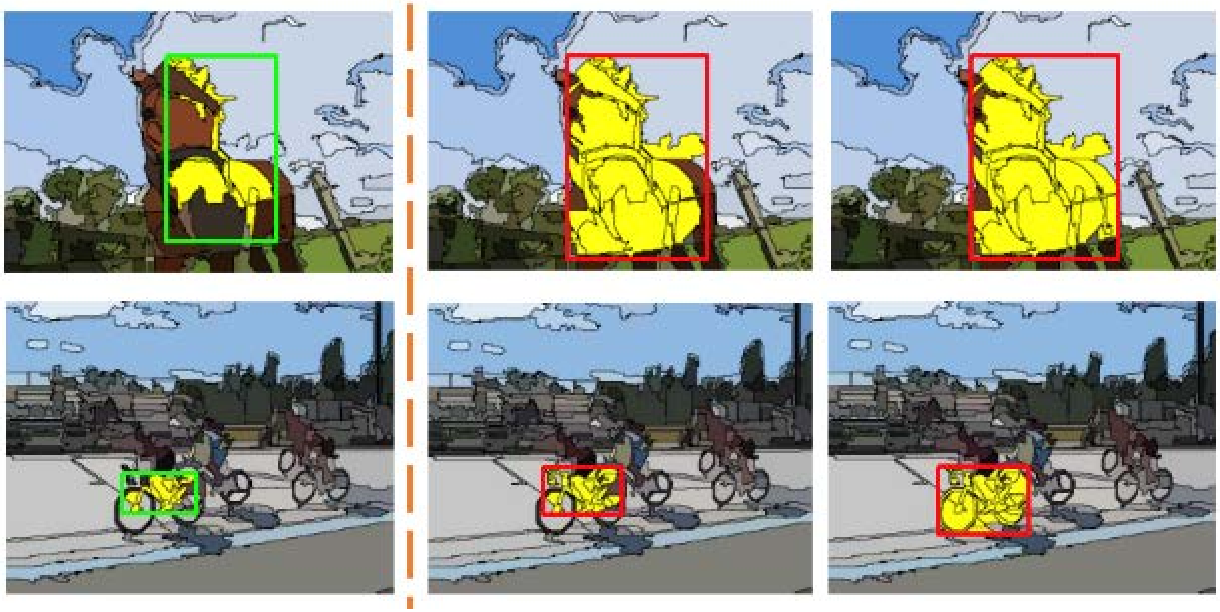
- Bounding box based approaches
  - Faster
- Segmentation based approaches
  - More accurate localization

Recent work: add-ons

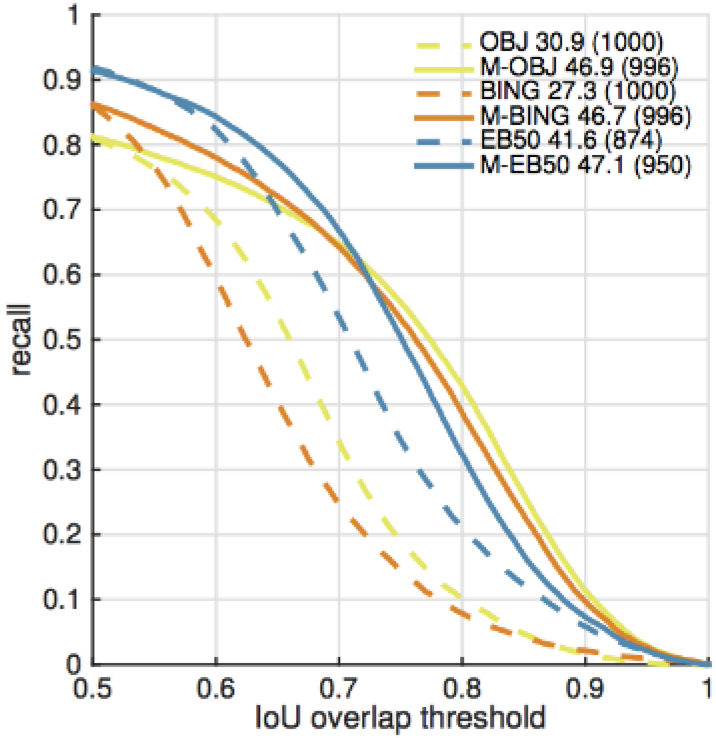
- Research on improve existing approaches
  - MTSE (CVPR'15)
  - Contour box (ICCV'15)

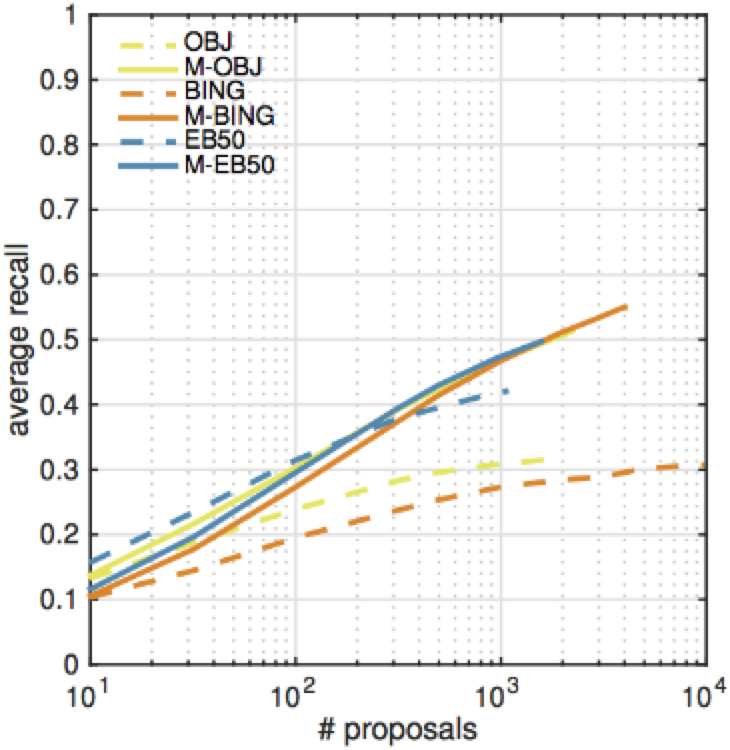
Multi-Thresholding Straddling Expansion (MTSE)

- Superpixel generation
- Expand bounding boxes to include straddling superpixels



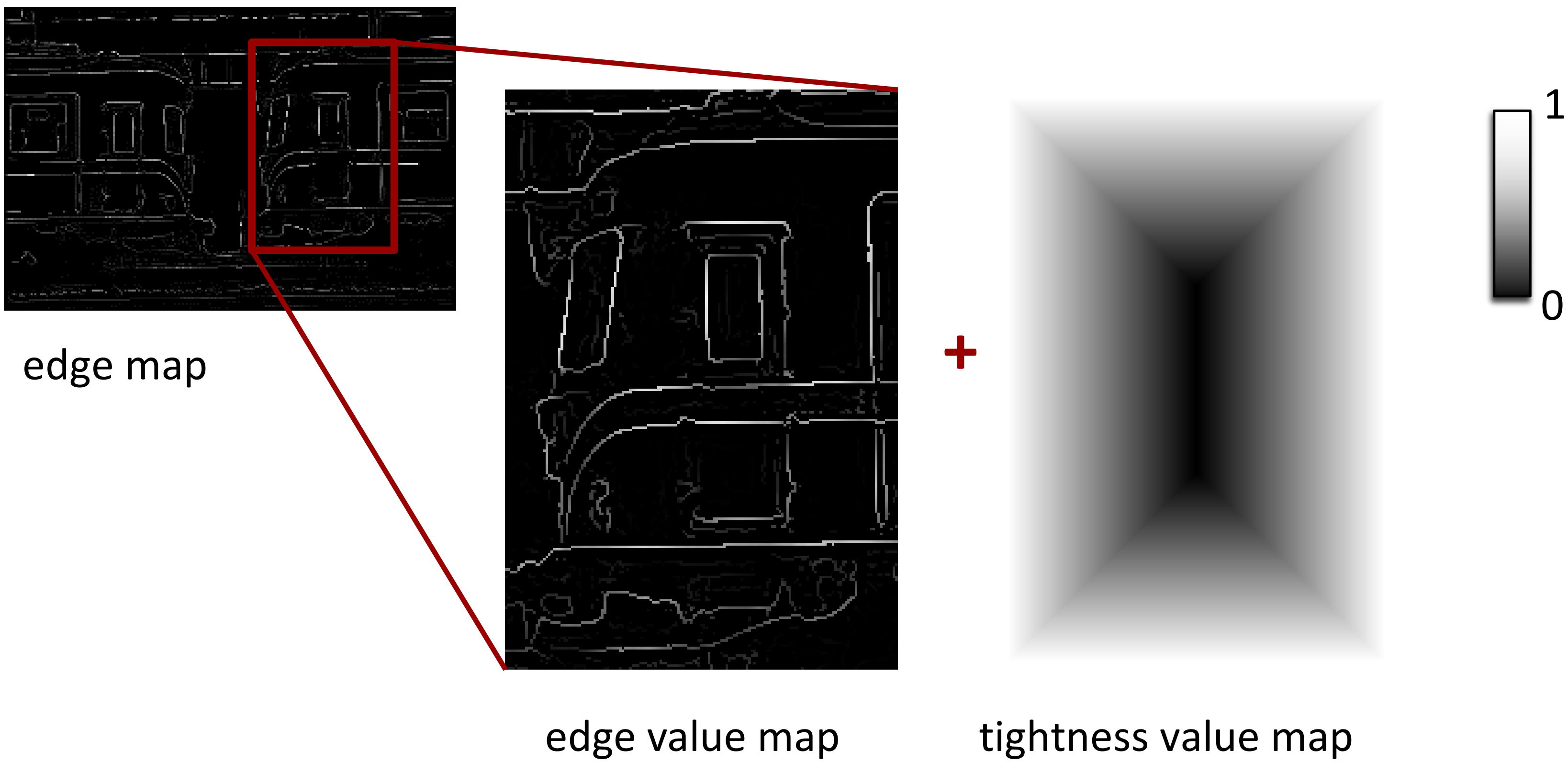
Multi-Thresholding Straddling Expansion (MTSE)





Contour box

- Edge detection (SE detector)
- Find the **optimal contour** in a box
  - Completeness -- penalty for opening
  - Tightness -- distance to the border

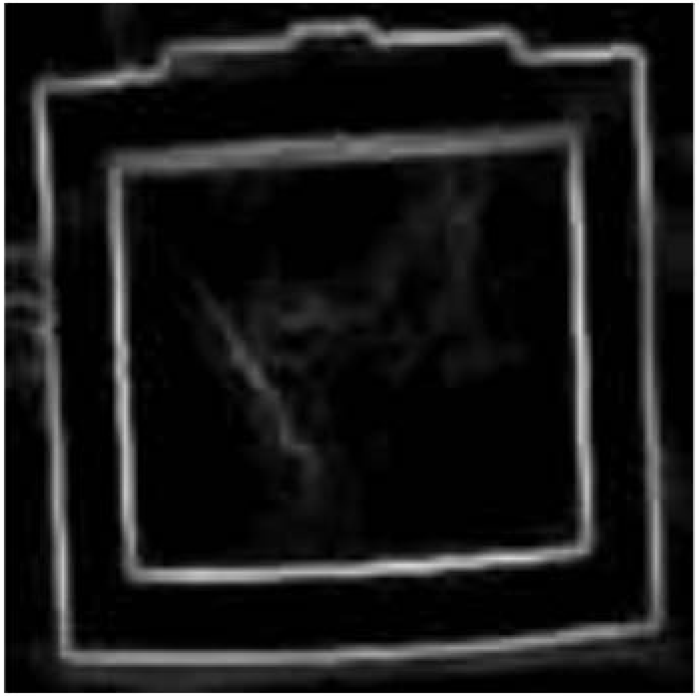


Goal: find a closed path that maximizes (sum of value / path length)

Contour box

- The optimal contour problem is **NP-hard** in Cartesian coordinates
- Solution: dynamic programming can give a good approximation in **polar coordinates**

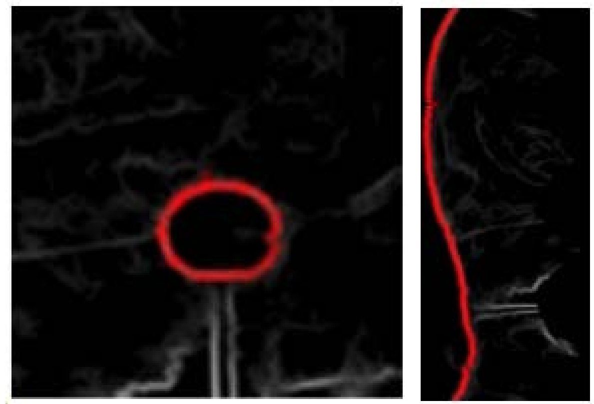
Contour box



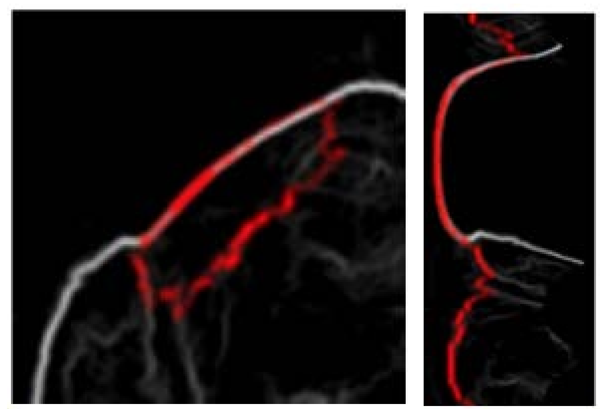
Goal: find path from top to down that maximize the sum of value
Easily solved by dynamic programming

Contour box

A box with high score



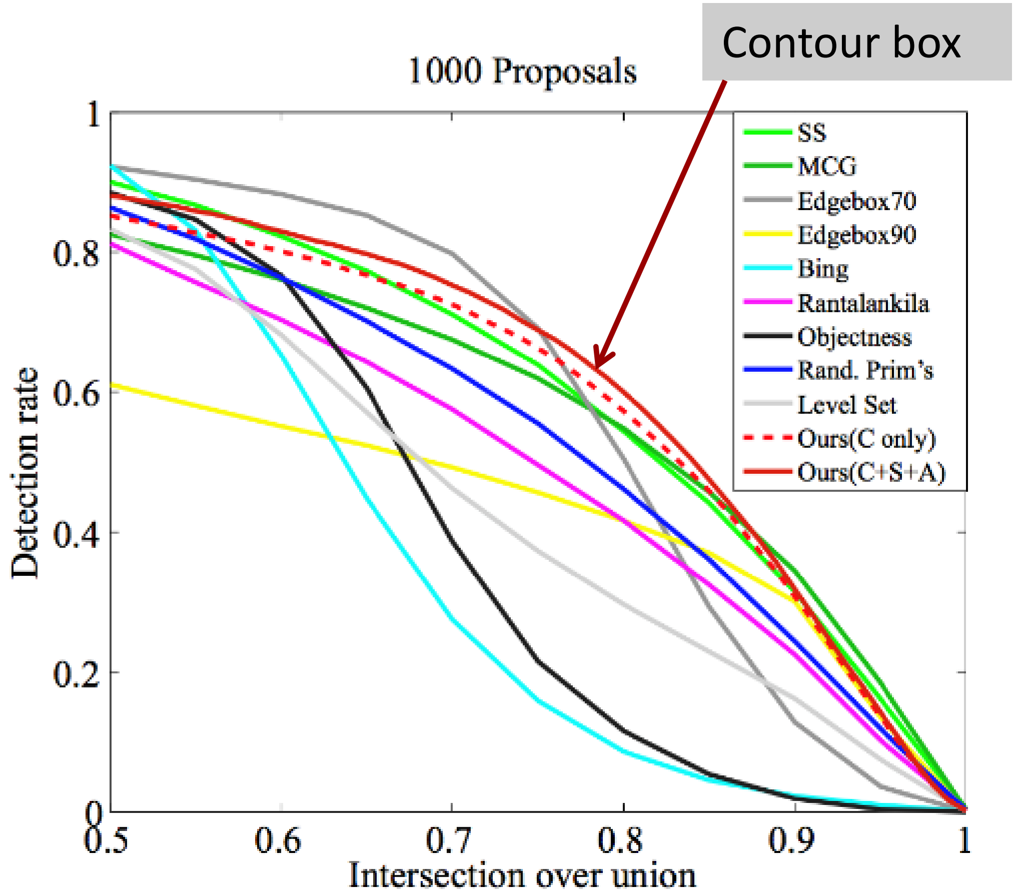
Not tight enough



Open contour

Boxes with low score

Contour box

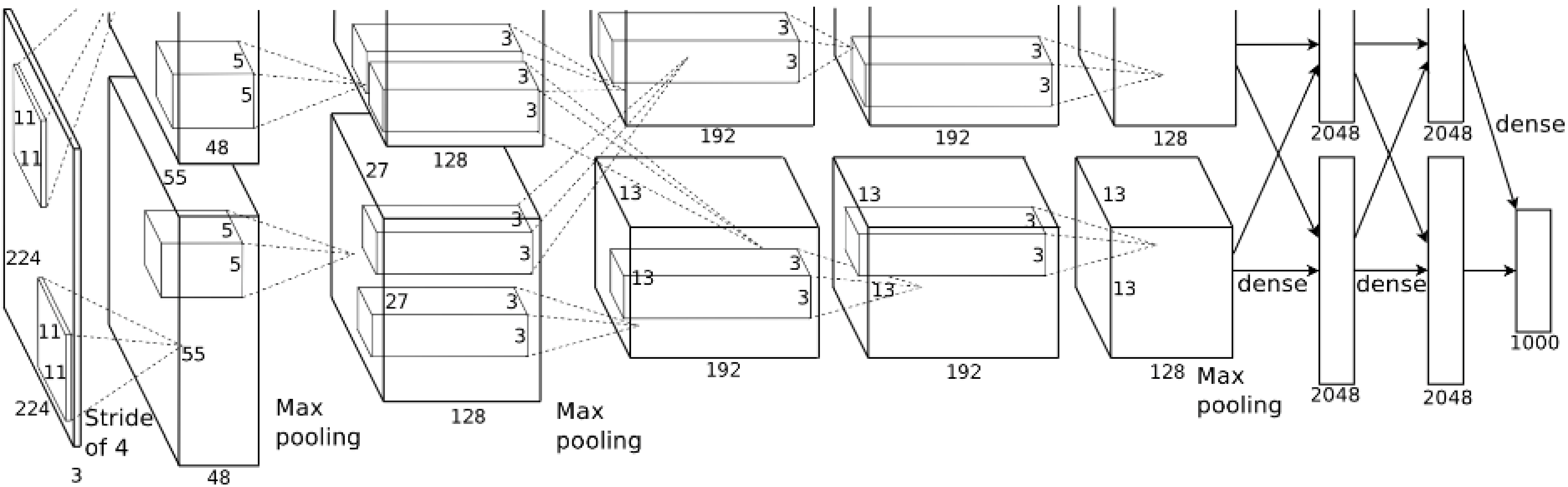

Contour box
1000 Proposals
Detection rate
Intersection over union
0
0.2
0.4
0.6
0.8
1
0.5
0.6
0.7
0.8
0.9
1
SS
MCG
Edgebox70
Edgebox90
Bing
Rantalankila
Objectness
Rand. Prim's
Level Set
Ours(C only)
Ours(C+S+A)

Recent work: CNN and object proposal

- Use CNN features for proposal
  - **Deep Proposal** (ICCV'15)
  - Novotny et al. (ICCV'15)
- Use CNN to re-rank proposals
  - **Deep Box** (ICCV'15)
- Share CNN features for both proposal and detection
  - **Faster R-CNN** (NIPS'15)

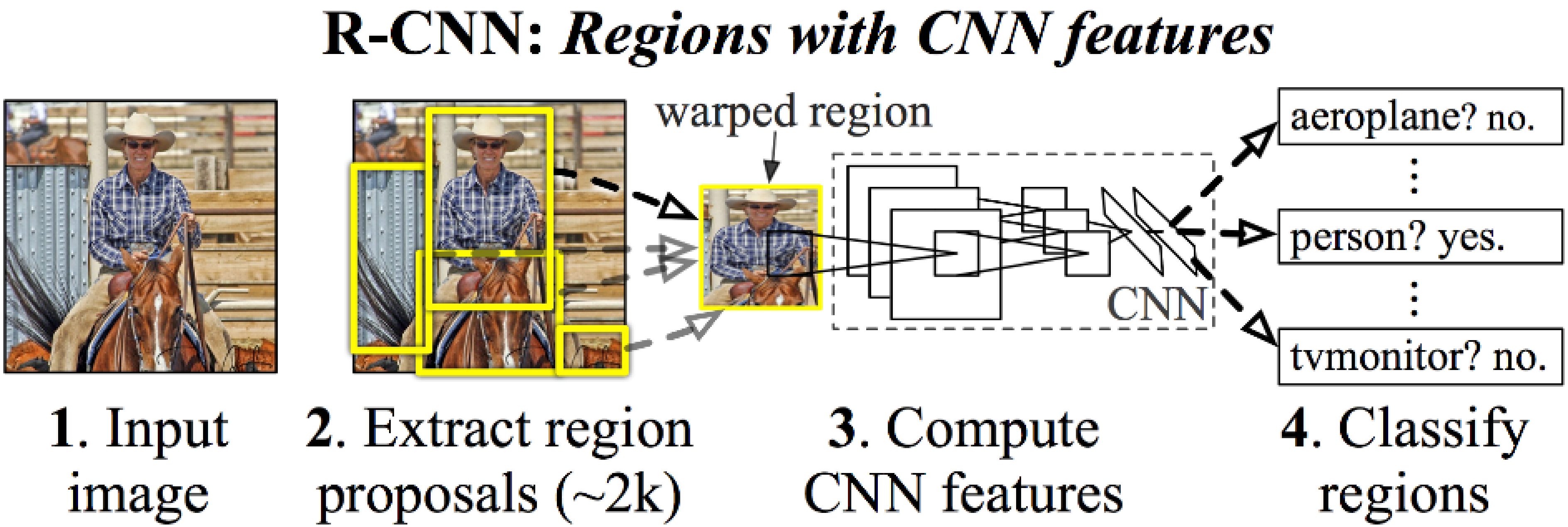
CNN brief review

- AlexNet (NIPS'12): Image classification



CNN and object proposal

- Earliest CNN detector: R-CNN (CVPR'14)

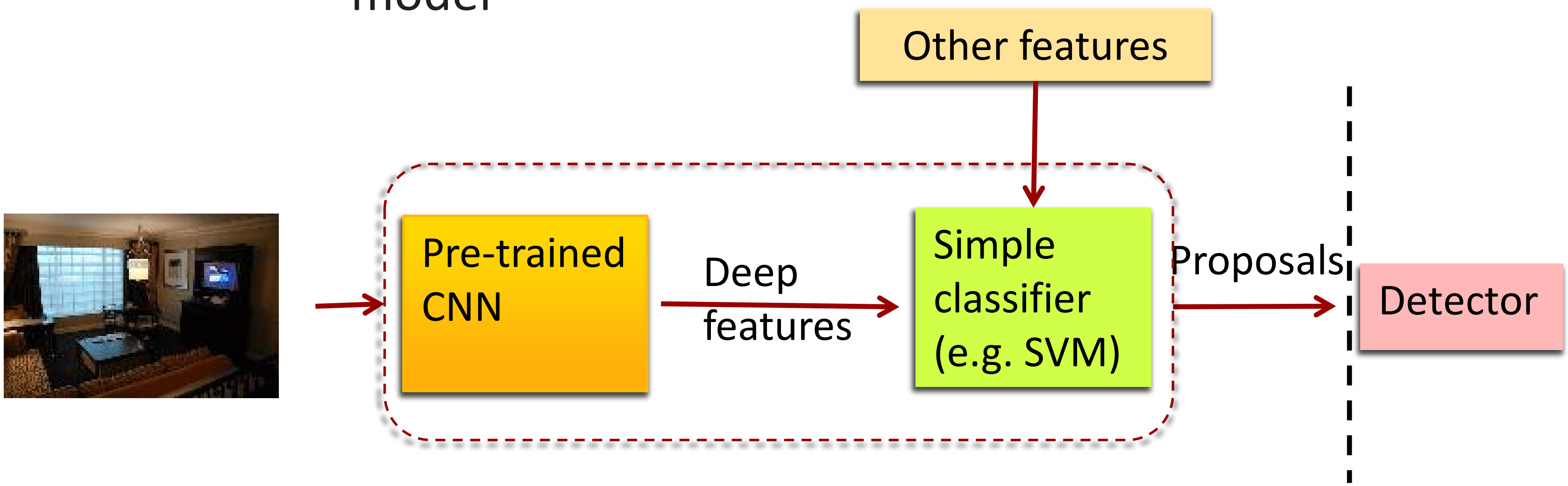


CNN and object proposal

- R-CNN uses object proposals to reduce computation load
  - CNN-based object detector is much more expensive than traditional detectors
  - Applying CNN-based detector on all possible regions (e.g. sliding window) is computationally impossible
- R-CNN largely encourages research in object proposal since it is important for CNN-based detectors
- Later, people leverage the power of CNN to generate object proposals

Deep Proposal

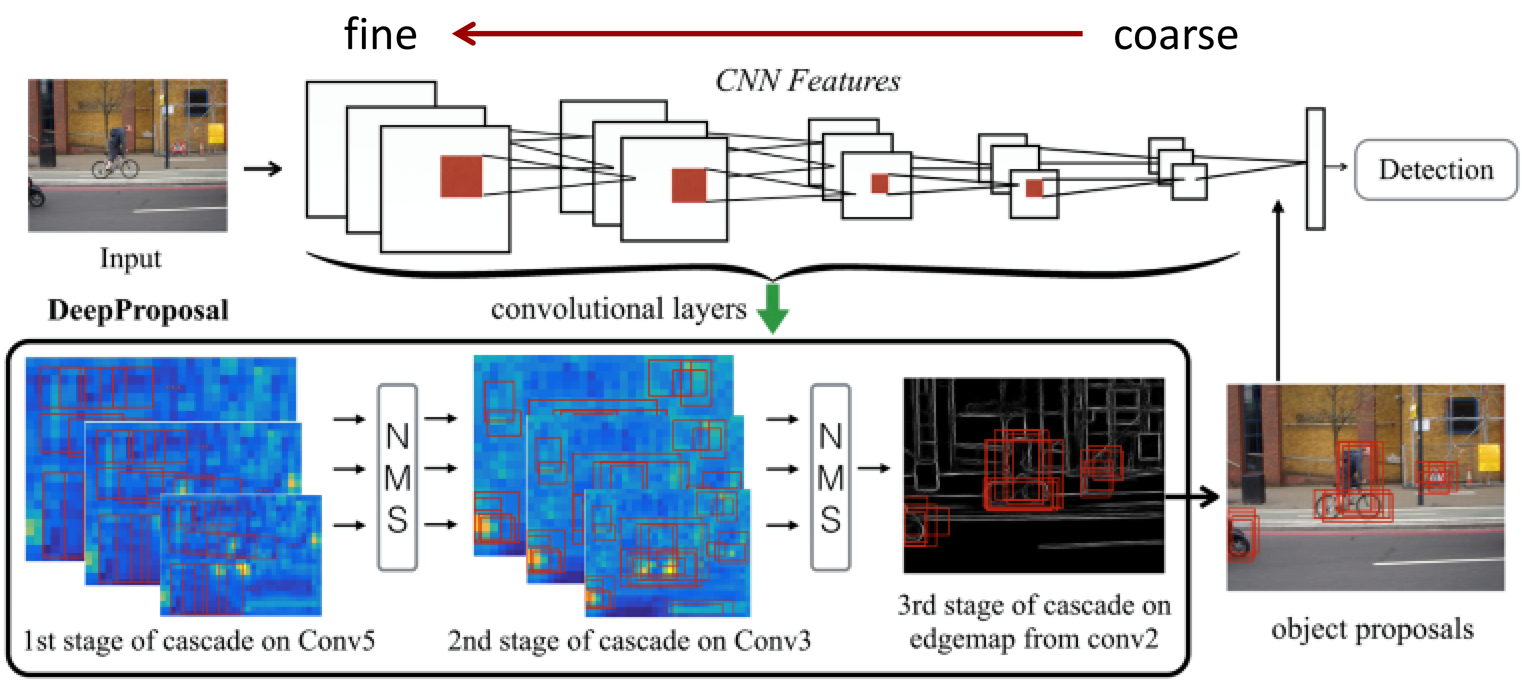
- Train a simple classifier on CNN features (plus non-CNN features)
  - CNN features are extracted from a pre-trained model



Ghodrati, Amir, et al. "DeepProposal: Hunting Objects by Cascading Deep Convolutional Layers." ICCV 2015

Deep Proposal

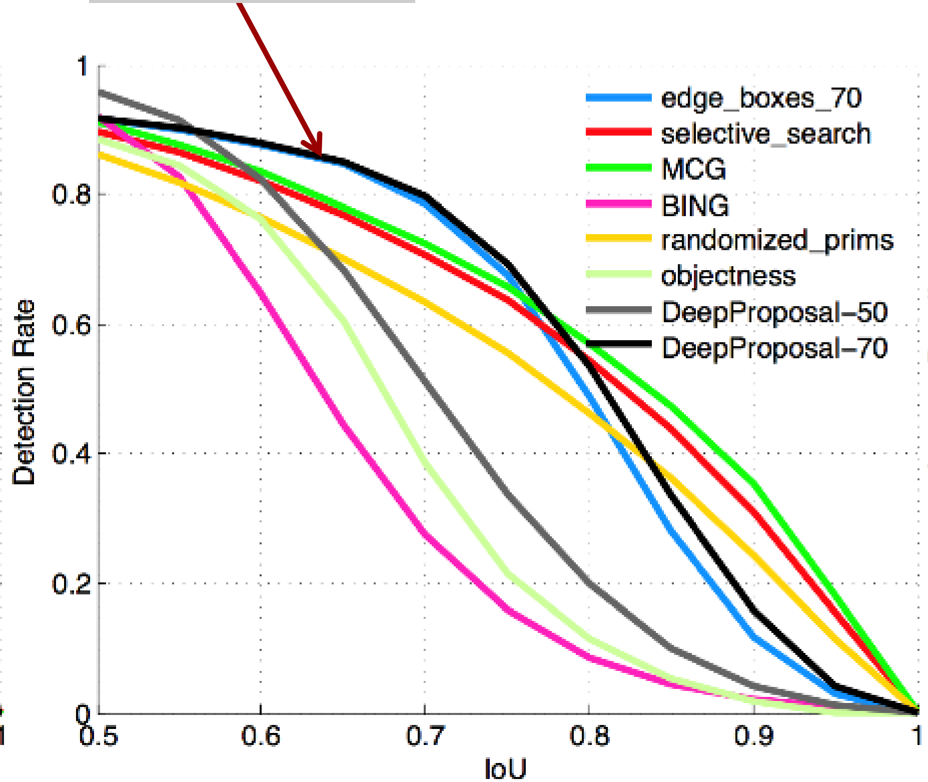
- Reversely cascade conv layers
  - High conv layers have summary of a region
  - Low conv layers have location details

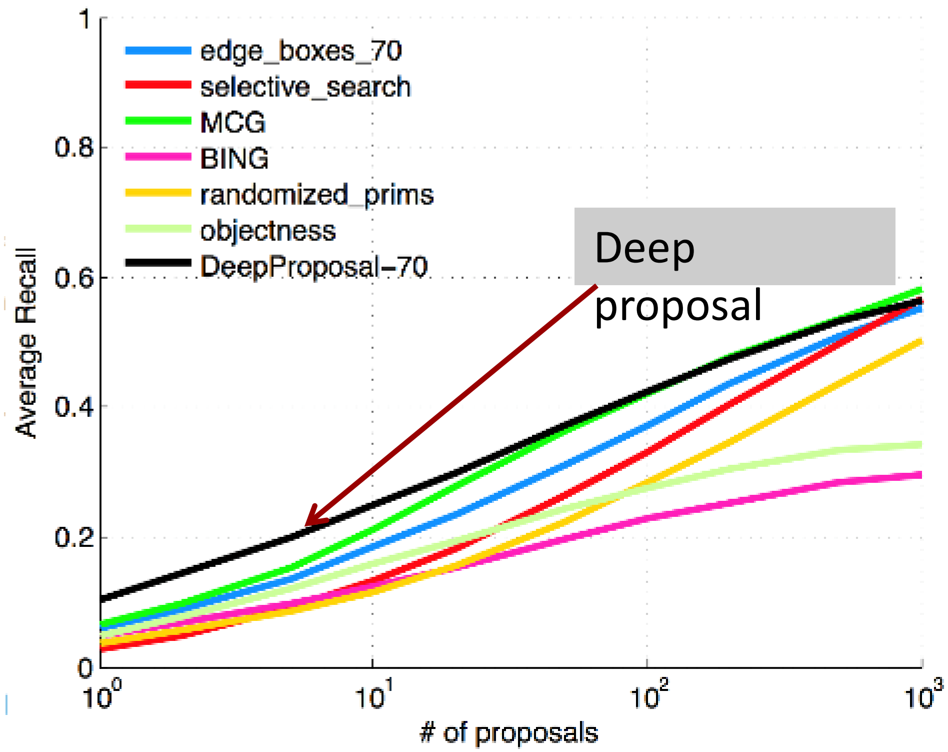


Deep Proposal

- Reduce the number of candidates

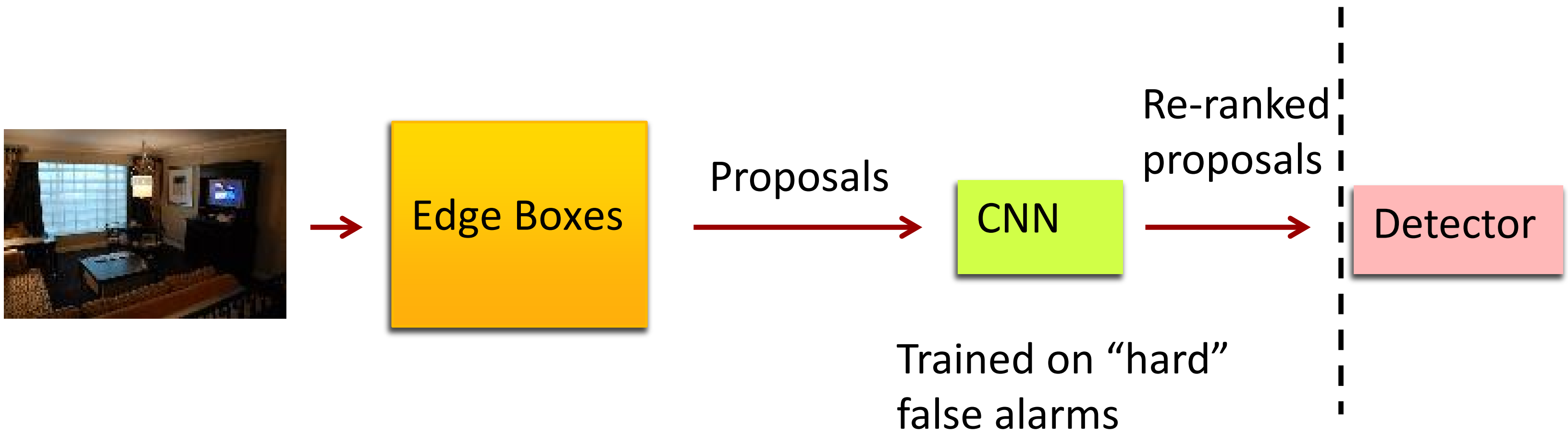
Deep proposal





Deep Box

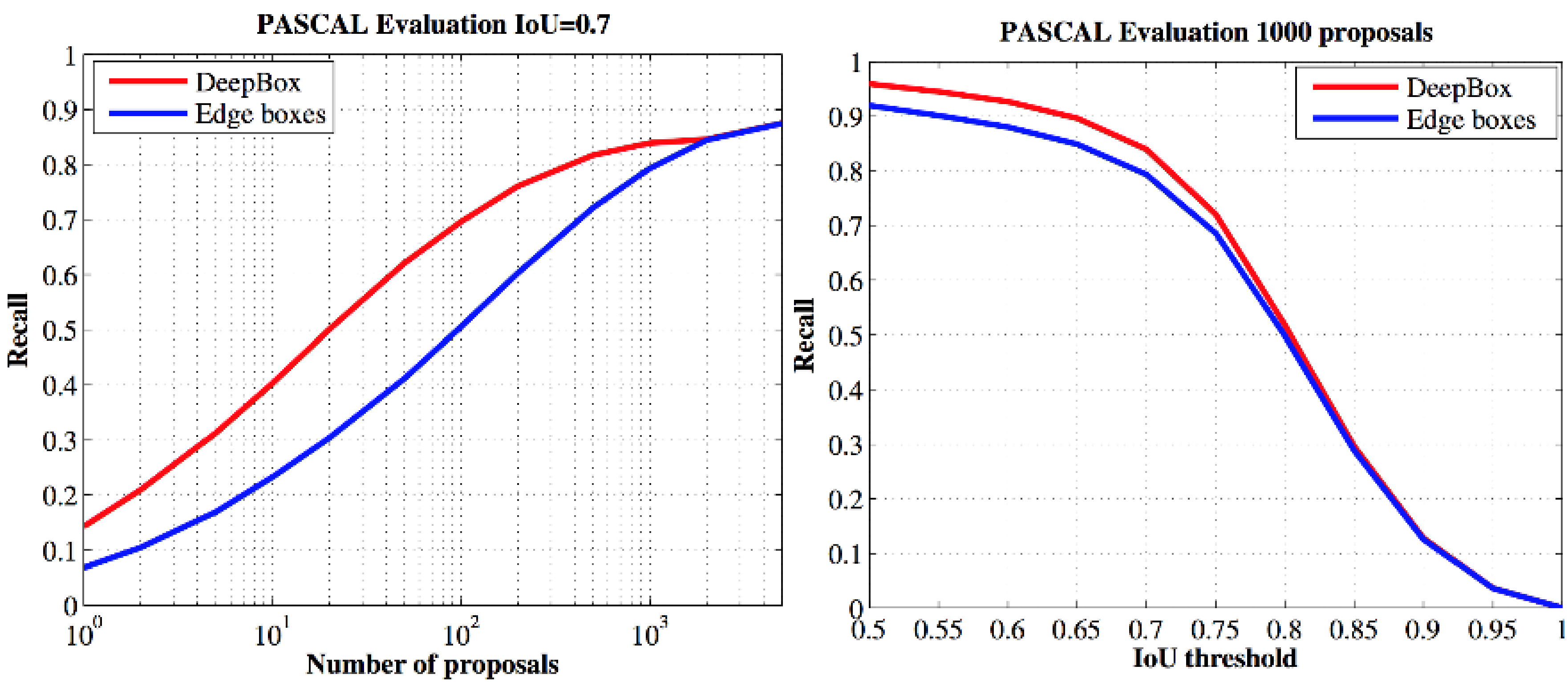
- Re-rank proposals from non-CNN approaches



Kuo Weicheng, Bharath Hariharan, Jitendra Malik. "DeepBox: Learning Objectness with Convolutional Networks." ICCV 2015

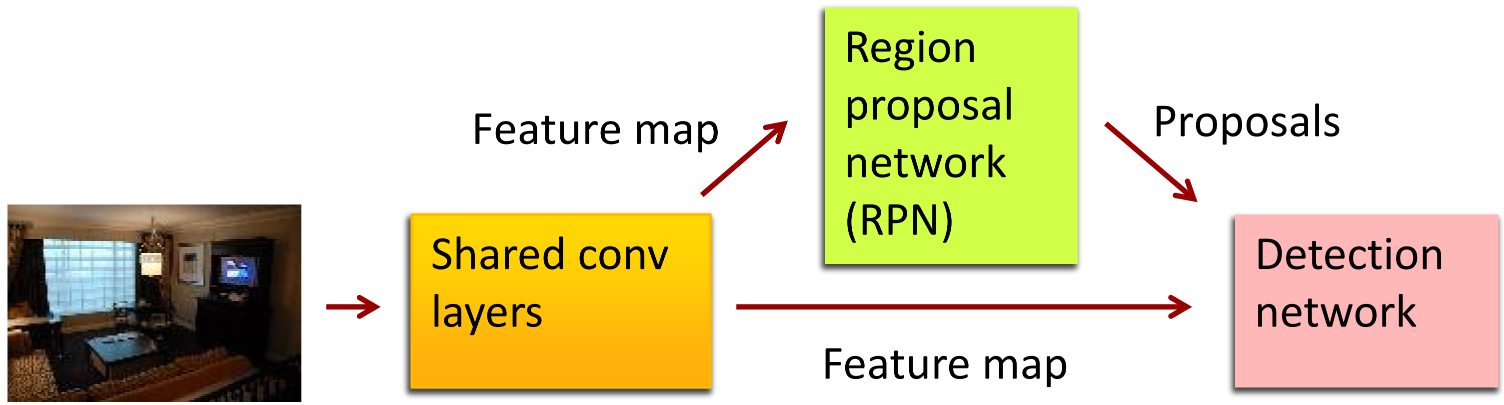
Deep Box

- Higher recall with less proposals
- Tuned for specific proposal approaches



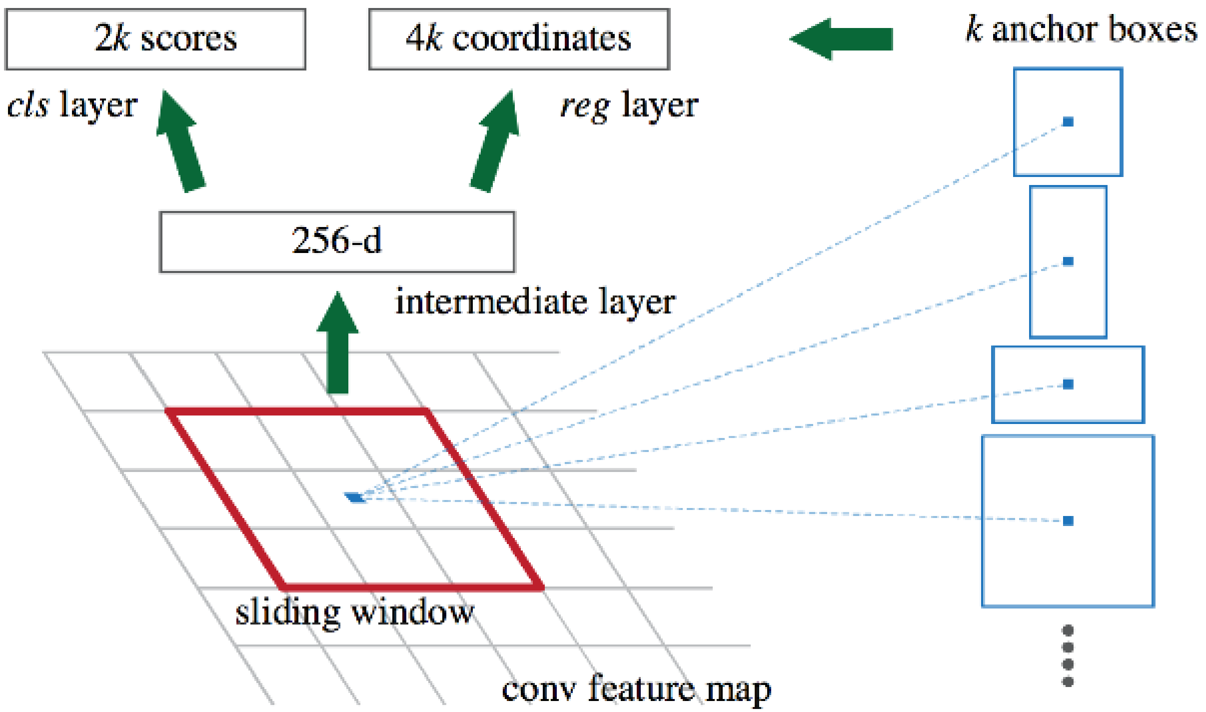
Faster R-CNN

- Share conv layers between proposal network and detection network
  - RPN is tuned for specific object categories and detection classifiers



Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks”. NIPS 2015.
Szegedy, Christian, et al. "Going deeper with convolutions." arXiv preprint arXiv:1409.4842 (2014).

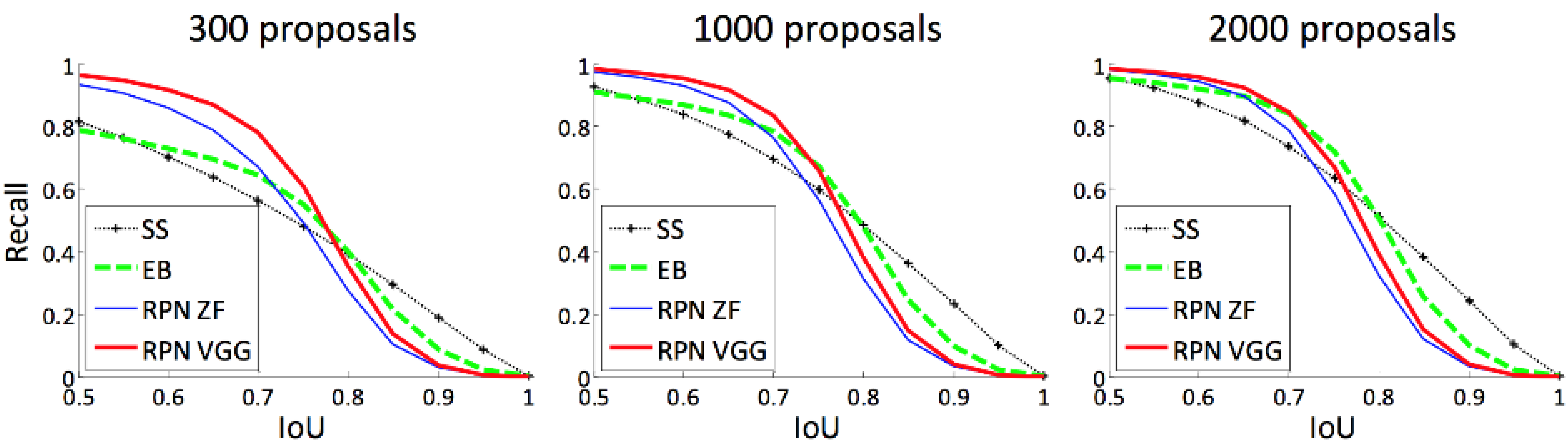
Faster R-CNN



Region Proposal Network: a module to adjust bounding boxes based on anchor boxes

Faster R-CNN

- Highly reduces the number of candidates needed for detection because of fine-tuning



Future research

- Generalization across dataset
  - The size of dataset grows; fully-labeled dataset not possible
  - It is extremely hard to label <u>all objects</u> in an image

Future research

- Combine with CNN detectors
  - What property is the most important for CNN detectors?
    - Which one is better? Bounding box better or segmentation?
    - Does CNN need really accurate localization?

Future research

- Computation efficiency
  - Consider advantages of GPU