DALHOUSIE UNIVERSITY

Faculty of Computer Science
Dalhousie University

Object localization by optimizing convolutional neural network detection score using generic edge features

Elham Etemad
Qigang Gao

September 15, 2017

Content

Outline

Introduction

- **Object**: Area in the image whose visual characteristics is learned by the computer
- **Object Detection**: Existence of a single object in the image
- **Object Localization**: Finding the accurate location of the detected object
- **Object Recognition**: Localizing all the presented objects
- **Scene Understanding**: Recognizing all objects and finding their roles
- Object Recognition is **essential technique** in computer vision based applications

Introduction

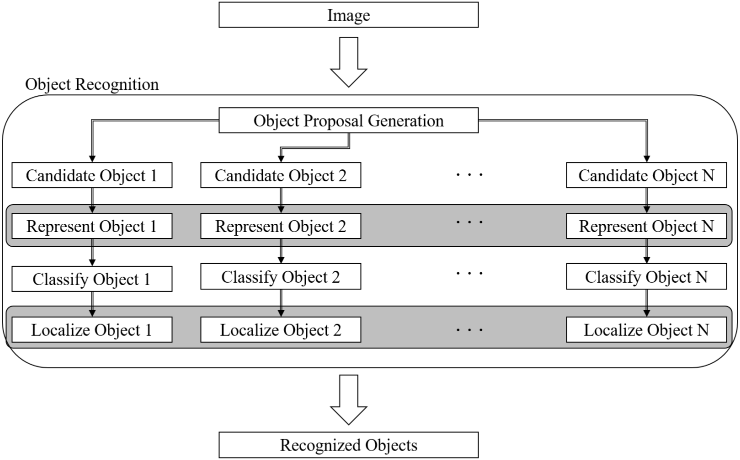


Figure: The main pipeline for many object recognition methods

Outline

Object Proposal Generation

- Sliding Window
- Selective Search
- Multi-branch Hierarchical Segmentation
- Complexity Adaptive Distance Metric
- Learning to Segment using RNN

Local Image Representation

- **Keypoint Detection**: SIFT, SURF, ORB, BRISK, FAST
- **Feature Description**: SIFT, SURF, ORB, BRISK, BRIEF, FREAK
- **Image Encoding**: Vector Quantization, Sparse Coding (SC), LLC, Group SC, Automatic Group SC, Label Constraint SC

Global Image Representation

- Color, Texture, Shape

Deep Image Representation

- Alex-Net, ZF-Net, VGG-Net

Combined Image Representation

- Local+Global, Local+Deep, Global+Deep, Local+Global+Deep

Object Localization

- Super-pixel Tightness
- Multiple Instance Learning
- Kernel Ridge Regressors

Object Recognition

- R-CNN
- Fast R-CNN
- DeepID-Net

Outline

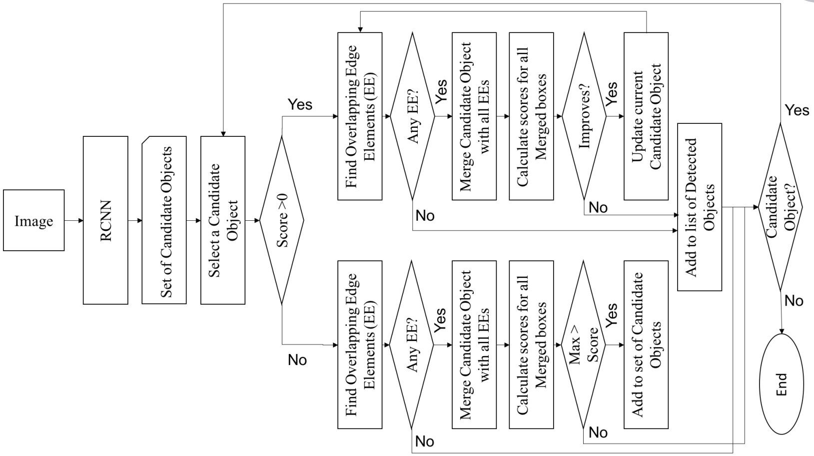


Figure: Main diagram of the proposed object localization method.

Object Localization
Proposed Method

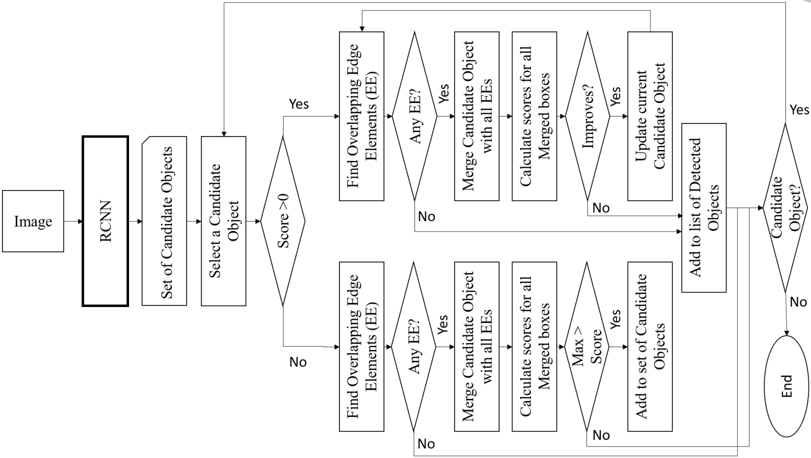


Figure: Main diagram of the proposed object localization method.

Candidate Object Detection

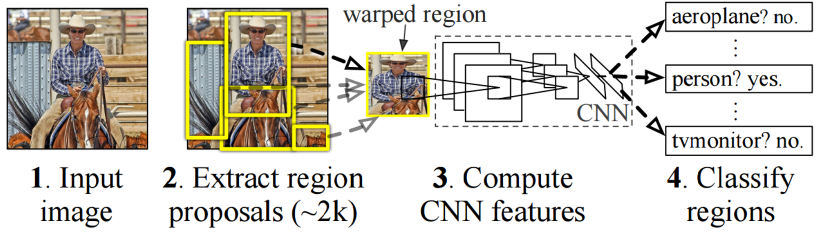


Figure: Main diagram of RCNN Object Recognition Module [1].

Object Localization
Proposed Method

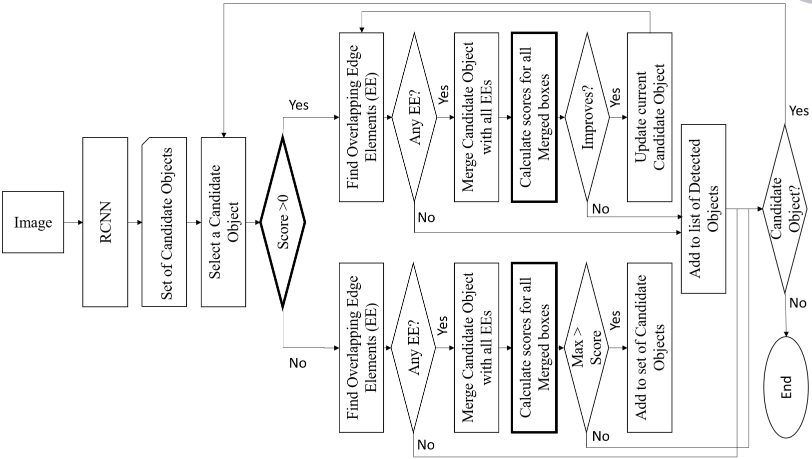
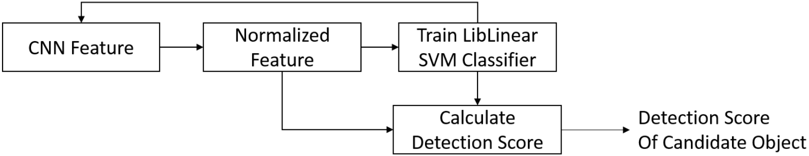


Figure: Main diagram of the proposed object localization method.

Detection Score

- **NormalizedFeature** $= C \times \frac{Feature}{\frac{1}{N} \times \sum_T Feature}$
- **Train Classifier**: $w = \min_{\acute{w}} \sum_{(f,l) \in T} \ell(\acute{w}; (f, l)) + Kr(\acute{w})$
- **Find Detection Score**: $\varphi(A, T) = f(A) \times w(T) + b(T)$



Object Localization

Proposed Method

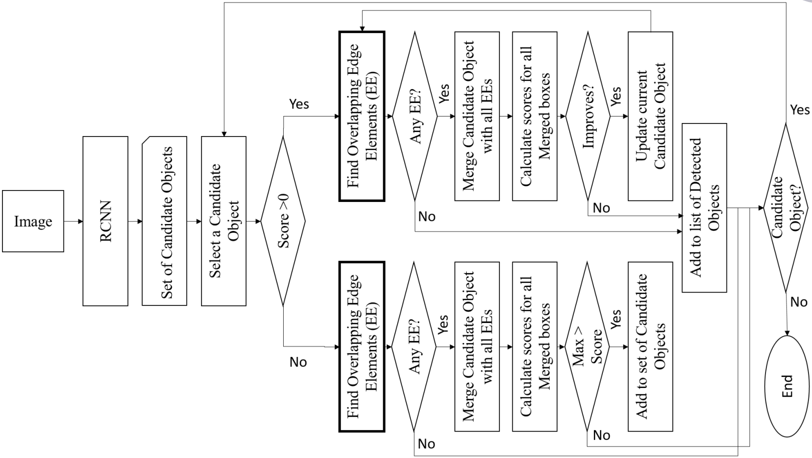
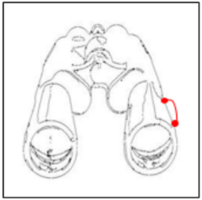
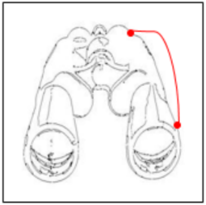


Figure: Main diagram of the proposed object localization method.

Edge Elements


(a) Image


(b) GET


(c) Trace

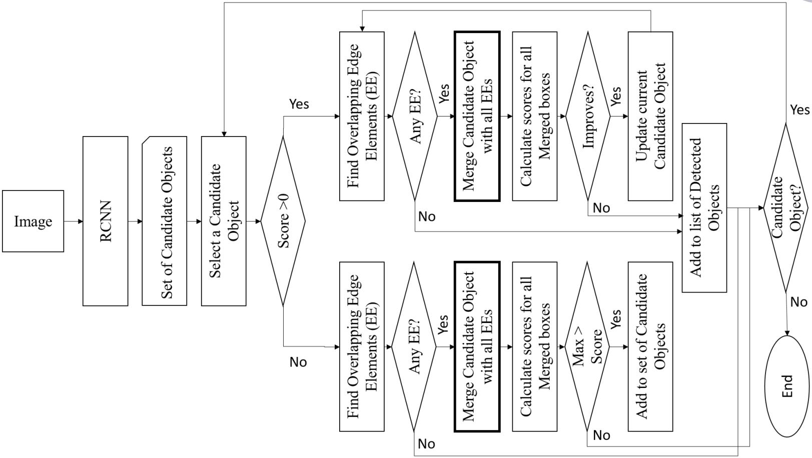
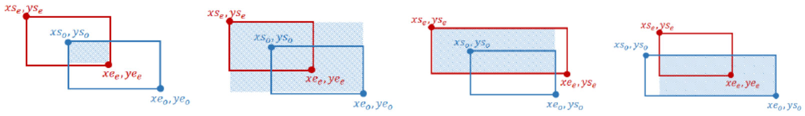


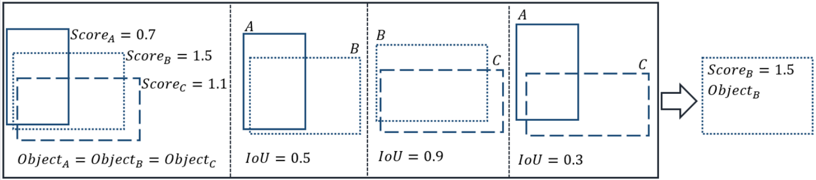
Figure: Main diagram of the proposed object localization method.

Object Localization
Proposed Method

Merge Bounding Boxes



Non-Maximum Suppression



Optimization Algorithm

Algorithm 6 Object localization using the Generic Edge Tokens of the image

procedure **GETLoc**($Image, CanObj$)
    ▷ *Input: Image*
    ▷ *Input: List of candidate boxes with their detection scores*
    ▷ *Output: List of detected boxes with their detection scores*
    for *Each* $CandidBox_i$ **do**
        while *Detection Score Improves* **do**
            $FindMergedBoxes(CandidBox, EdgeMap)$
            for *Each Merged Box* j **do**
                ▷ *Calculate Detection Score* $DS_{i,j}$
                $DS_{i,j} = CNNScore(MergedBox_j)$
            ▷ *Find the best merged box*
            $SelectedBox = arg\ max_{j \in MergedBox} DS_{i,j}$
            $CandidBox_i = SelectedBox$

Optimization Iterations

Iter = 0, S=-0.18, IoU = 0.47

Iter = 1, S=0.25, IoU = 0.54

Iter = 2, S=0.89, IoU = 0.58

Iter = 3, S=2.19, IoU = 0.66

Iter = 4, S=3.10, IoU = 0.70

Iter = 5, S=3.26, IoU = 0.76

Figure: Improved bounding boxes after several iterations.

Experimental Framework

- **Datasets**:
  - PASCAL VOC 2007
    - 20 classes, 9,963 images, 24,640 annotated objects
    - **test** set, 4952 images
    - **validation** set, 2510 images
  - PASCAL VOC 2012
    - 20 classes, 22,521 images, 27,450 annotated objects in training set
    - **test** set, 10991 images
- **Measurements**
  - $AP = \frac{number \quad of \quad detected \quad objects}{total \quad number \quad of \quad objects}$
  - $mAP = \frac{\sum_N AP}{N}, \quad N = number \quad of \quad classes$
- **Packages**
  - RCNN using AlexNet
  - Caffe
  - PCPG

Object Localization
Experimental Results

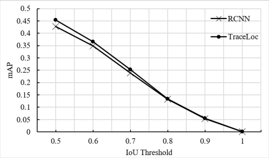
Class based and Global mAP

(a) Test 2007	Aero	Bike	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow	Table	Dog	Horse	Mbike	Person	Plant	Sheep	Sofa	Train	TV	mAP
RCNN	49.8	**61.7**	32.8	25.2	24.2	**53.1**	61.5	49.0	22.8	48.8	33.2	39.4	51.4	51.5	48.4	15.6	**50.2**	35.0	**49.5**	51.2	42.7
GET_Loc	49.5	60.9	37.7	31.0	30.3	51.2	61.4	**54.4**	27.8	**53.7**	32.6	46.1	57.5	58.4	48.5	**20.8**	48.2	34.1	47.9	51.6	45.2
Trace_Loc	50.3	61.3	**39.8**	31.6	30.8	51.9	61.9	48.6	28.9	47.6	**34.3**	**47.1**	**58.7**	59.6	48.5	20.6	49.3	**35.5**	49.0	**52.2**	45.4
GT_Loc	**50.4**	61.3	39.4	**31.8**	**32.0**	52.3	**62.0**	48.9	**29.2**	47.8	33.3	46.8	58.4	**59.6**	**48.6**	20.7	48.4	35.4	49.2	51.9	**45.4**
(b) Test 2012	Aero	Bike	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow	Table	Dog	Horse	Mbike	Person	Plant	Sheep	Sofa	Train	TV	mAP
RCNN	56.4	49.3	31.4	15.4	19.4	43.3	46.1	52.4	13.6	31.9	23.8	48.7	41.1	51.8	44.0	12.8	42.9	20.4	33.7	34.4	35.6
GET_Loc	**59.2**	52.7	**35.5**	**18.8**	22.7	46.0	49.0	55.1	17.2	**38.1**	26.4	51.3	**44.5**	53.8	47.0	**14.9**	44.7	**23.3**	**38.3**	**39.1**	**38.9**
Trace_Loc	58.4	**53.3**	35.2	**18.8**	22.5	46.5	48.6	54.9	16.6	37.8	25.8	**51.9**	43.7	**54.5**	**47.3**	13.8	44.3	22.2	37.8	38.4	38.6
GT_Loc	58.8	52.8	35.0	18.7	**23.1**	**46.8**	**49.1**	**55.2**	**17.5**	37.8	**26.5**	51.4	44.4	54.1	47.1	14.7	**45.3**	23.1	**38.3**	**39.1**	**38.9**
(c) Validation 2007	Aero	Bike	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow	Table	Dog	Horse	Mbike	Person	Plant	Sheep	Sofa	Train	TV	mAP
RCNN	**81.1**	80.1	70.2	53.7	43.0	71.2	**71.3**	**80.1**	61.5	**81.3**	62.7	**81.1**	81.6	80.5	50.7	33.8	**70.8**	72.7	**81.5**	72.6	69.1
GET_Loc	80.0	80.2	79.0	70.0	47.8	71.0	71.0	79.1	**66.9**	81.3	71.4	80.0	80.1	79.5	**56.9**	**41.7**	69.4	80.3	79.7	**81.4**	72.3
Trace_Loc	79.5	80.1	69.7	**70.4**	**50.8**	**71.3**	70.7	79.5	58.8	80.7	**71.6**	79.6	80.5	79.6	56.4	40.5	70.3	**80.7**	79.5	81.2	71.6
GT_Loc	79.6	**80.5**	**79.5**	70.2	48.6	71.1	71.0	79.2	60.0	80.7	71.4	80.1	**88.3**	**87.6**	56.5	40.9	70.0	80.3	79.2	81.2	**72.8**

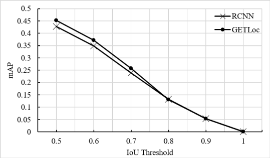
Object Localization

Experimental Results

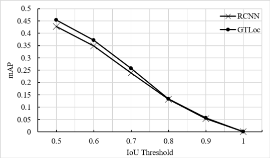
mAP vs **IoU** for VOC 2007 Test (a,b,c) and Validation (d,e,f) sets



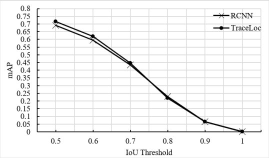
(a)



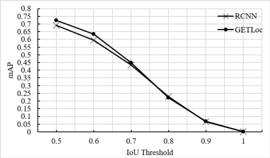
(b)



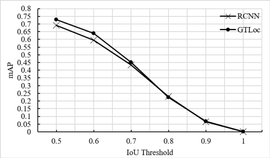
(c)



(d)



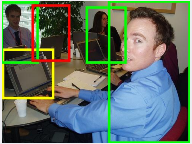
(e)



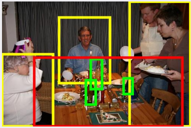
(f)

Object Localization
Experimental Results

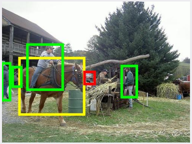
Samples of images from PASCAL VOC 2007 test set



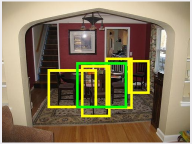
Yellow: Monitor
Green: Person
Red: Plant



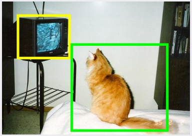
Yellow: Person
Green: Bottle
Red: Table



Yellow: Horse
Green: Person
Red: Car

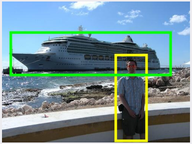


Yellow: Chair
Green: Table



Yellow: Monitor
Green: Cat

Yellow: Sofa
Green: Person



Yellow: Person
Green: Boat

Yellow: Aeroplane
Green: Person

Outline

Conclusion

Future Work

- Improving object localization by using a combination of the image edge, color and texture information, and the learned features of the image
- Proposing a way to have a non greedy suppression of the detected bounding boxes
- Proposing better object representation methods that considers the entire image context

Bibliography

[1] Ross Girshick, Je Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580-587, 2014.