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

Elham Etemad
Qigang Gao

September 15, 2017
Content

- Introduction

- Literature Review
  - Object Proposal Generation
  - Image Representation
  - Object Localization
  - Object Recognition

- Object Localization by Optimizing Convolutional Neural Network Detection Score using Generic Edge Features
  - Proposed Method
  - Experimental Results

- Conclusion
Outline

Introduction

Literature Review
- Object Proposal Generation
- Image Representation
- Object Localization
- Object Recognition

Object Localization by Optimizing Convolutional Neural Network Detection Score using Generic Edge Features
- Proposed Method
- Experimental Results

Conclusion
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

- Introduction

- Literature Review
  - Object Proposal Generation
  - Image Representation
  - Object Localization
  - Object Recognition

- Object Localization by Optimizing Convolutional Neural Network Detection Score using Generic Edge Features
  - Proposed Method
  - Experimental Results

- Conclusion
Object Proposal Generation

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

#### Image Representation

<table>
<thead>
<tr>
<th>Local Image Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Keypoint Detection:</strong> SIFT, SURF, ORB, BRISK, FAST</td>
</tr>
<tr>
<td><strong>Feature Description:</strong> SIFT, SURF, ORB, BRISK, BRIEF, FREAK</td>
</tr>
<tr>
<td><strong>Image Encoding:</strong> Vector Quantization, Sparse Coding (SC), LLC, Group SC, Automatic Group SC, Label Constraint SC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Global Image Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Color, Texture, Shape</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deep Image Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alex-Net, ZF-Net, VGG-Net</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Combined Image Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Local+Global, Local+Deep, Global+Deep, Local+Global+Deep</strong></td>
</tr>
</tbody>
</table>
Literature Review
Object Localization and Recognition

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

Object Recognition
- R-CNN
- Fast R-CNN
- DeepID-Net
Outline

- Introduction

- Literature Review
  - Object Proposal Generation
  - Image Representation
  - Object Localization
  - Object Recognition

- Object Localization by Optimizing Convolutional Neural Network Detection Score using Generic Edge Features
  - Proposed Method
  - Experimental Results

- Conclusion
Figure: Main diagram of the proposed object localization method.
Figure: Main diagram of the proposed object localization method.
Candidate Object Detection

1. Input image
2. Extract region proposals (~2k)
3. Compute CNN features
4. Classify regions

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

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

Figure: Main diagram of the proposed object localization method.
Detection Score

- **Normalized Feature**: \[ \text{Normalized Feature} = C \times \frac{\sum_{T} \text{Feature}}{N} \]
- **Train Classifier**: \[ w = \min_{\hat{w}} \sum_{(f,l) \in T} \ell(\hat{w}; (f, l)) + Kr(\hat{w}) \]
- **Find Detection Score**: \[ \varphi(A, T) = f(A) \times w(T) + b(T) \]
Figure: Main diagram of the proposed object localization method.
Object Localization
Proposed Method

Edge Elements

(a) Image  (b) GET  (c) Trace

Elham Etemad | Object localization by optimizing convolutional neural network detection score using generic edge features
Figure: Main diagram of the proposed object localization method.
Object Localization
Proposed Method

Merge Bounding Boxes

Elham Etemad | Object localization by optimizing convolutional neural network detection score using generic edge features
Object Localization
Proposed Method

Non-Maximum Suppression

\[ \text{Score}_A = 0.7 \]
\[ \text{Score}_B = 1.5 \]
\[ \text{Score}_C = 1.1 \]

Object_A = Object_B = Object_C

\[ \text{IoU} = 0.5 \]
\[ \text{IoU} = 0.9 \]
\[ \text{IoU} = 0.3 \]

Score_B = 1.5
Object_B
ObjectLocalization
ProposedMethod

OptimizationAlgorithm

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

1: procedure GETLoc(Image, CanObj)
2: \textgreater{} Input: Image
3: \textgreater{} Input: List of candidate boxes with their detection scores
4: \textgreater{} Output: List of detected boxes with their detection scores
5: for Each CandidBox$_i$ do
6: \textbf{while} Detection Score Improves \textbf{do}
7: \hspace{1em} FindMergedBoxes(CandidBox$_i$, EdgeMap)
8: \hspace{1em} for Each Merged Box j do
9: \hspace{2em} \textgreater{} Calculate Detection Score $DS_{i,j}$
10: \hspace{2em} $DS_{i,j} = \text{CNNScore(MergedBox}_j)$
11: \hspace{2em} \textgreater{} Find the best merged box
12: \hspace{1em} SelectedBox = $\arg \max_{j \in \text{MergedBox}} DS_{i,j}$
13: \hspace{1em} CandidBox$_i$ = SelectedBox
Object Localization
Proposed Method

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.
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:
- \( AP = \frac{\text{number of detected objects}}{\text{total number of objects}} \)
- \( mAP = \frac{\sum_{N} AP}{N}, \quad N = \text{number of classes} \)

Packages:
- RCNN using AlexNet
- Caffe
- PCPG
## Object Localization

### Experimental Results

### Class based and Global mAP

<table>
<thead>
<tr>
<th>(a) Test 2007</th>
<th>Aero</th>
<th>Bike</th>
<th>Bird</th>
<th>Boat</th>
<th>Bottle</th>
<th>Bus</th>
<th>Car</th>
<th>Cat</th>
<th>Chair</th>
<th>Cow</th>
<th>Table</th>
<th>Dog</th>
<th>Horse</th>
<th>Mbike</th>
<th>Person</th>
<th>Plant</th>
<th>Sheep</th>
<th>Sofa</th>
<th>Train</th>
<th>TV</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCNN</td>
<td>49.8</td>
<td>61.7</td>
<td>32.8</td>
<td>25.2</td>
<td>24.2</td>
<td>53.1</td>
<td>61.5</td>
<td>49.0</td>
<td>22.8</td>
<td>48.8</td>
<td>33.2</td>
<td>39.4</td>
<td>51.4</td>
<td>51.5</td>
<td>48.4</td>
<td>15.6</td>
<td>50.2</td>
<td>35.0</td>
<td>49.5</td>
<td>51.2</td>
<td>42.7</td>
</tr>
<tr>
<td>GET.Loc</td>
<td>49.5</td>
<td>60.9</td>
<td>37.7</td>
<td>31.0</td>
<td>30.3</td>
<td>51.2</td>
<td>61.4</td>
<td>54.4</td>
<td>27.8</td>
<td>53.7</td>
<td>32.6</td>
<td>46.1</td>
<td>57.5</td>
<td>58.4</td>
<td>48.5</td>
<td>20.8</td>
<td>48.2</td>
<td>34.1</td>
<td>47.9</td>
<td>51.6</td>
<td>45.2</td>
</tr>
<tr>
<td>Trace.Loc</td>
<td>50.3</td>
<td>61.3</td>
<td>39.8</td>
<td>31.6</td>
<td>30.8</td>
<td>51.9</td>
<td>61.9</td>
<td>48.6</td>
<td>28.9</td>
<td>47.6</td>
<td>34.3</td>
<td>47.1</td>
<td>58.7</td>
<td>59.6</td>
<td>48.5</td>
<td>20.6</td>
<td>49.3</td>
<td>35.5</td>
<td>49.0</td>
<td>52.2</td>
<td>45.4</td>
</tr>
<tr>
<td>GT.Loc</td>
<td>50.4</td>
<td>61.3</td>
<td>39.4</td>
<td>31.8</td>
<td>32.0</td>
<td>52.3</td>
<td>62.0</td>
<td>48.9</td>
<td>29.2</td>
<td>47.8</td>
<td>33.3</td>
<td>46.8</td>
<td>58.4</td>
<td>59.6</td>
<td>48.6</td>
<td>20.7</td>
<td>48.4</td>
<td>35.4</td>
<td>49.2</td>
<td>51.9</td>
<td>45.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) Test 2012</th>
<th>Aero</th>
<th>Bike</th>
<th>Bird</th>
<th>Boat</th>
<th>Bottle</th>
<th>Bus</th>
<th>Car</th>
<th>Cat</th>
<th>Chair</th>
<th>Cow</th>
<th>Table</th>
<th>Dog</th>
<th>Horse</th>
<th>Mbike</th>
<th>Person</th>
<th>Plant</th>
<th>Sheep</th>
<th>Sofa</th>
<th>Train</th>
<th>TV</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCNN</td>
<td>56.4</td>
<td>49.3</td>
<td>31.4</td>
<td>15.4</td>
<td>19.4</td>
<td>43.3</td>
<td>46.1</td>
<td>52.4</td>
<td>13.6</td>
<td>31.9</td>
<td>23.8</td>
<td>48.7</td>
<td>41.1</td>
<td>51.8</td>
<td>44.0</td>
<td>12.8</td>
<td>42.9</td>
<td>20.4</td>
<td>33.7</td>
<td>34.4</td>
<td>35.6</td>
</tr>
<tr>
<td>GET.Loc</td>
<td>59.2</td>
<td>52.7</td>
<td>35.5</td>
<td>18.8</td>
<td>22.7</td>
<td>46.0</td>
<td>49.0</td>
<td>55.1</td>
<td>17.2</td>
<td>38.1</td>
<td>26.4</td>
<td>51.3</td>
<td>44.5</td>
<td>53.3</td>
<td>47.0</td>
<td>14.9</td>
<td>44.7</td>
<td>23.3</td>
<td>38.3</td>
<td>39.1</td>
<td>38.9</td>
</tr>
<tr>
<td>Trace.Loc</td>
<td>58.4</td>
<td>53.3</td>
<td>35.2</td>
<td>18.8</td>
<td>22.5</td>
<td>46.5</td>
<td>48.6</td>
<td>54.9</td>
<td>16.6</td>
<td>37.8</td>
<td>25.8</td>
<td>51.9</td>
<td>43.7</td>
<td>54.5</td>
<td>47.3</td>
<td>13.8</td>
<td>44.3</td>
<td>22.2</td>
<td>37.8</td>
<td>38.4</td>
<td>38.6</td>
</tr>
<tr>
<td>GT.Loc</td>
<td>58.8</td>
<td>52.8</td>
<td>35.0</td>
<td>18.7</td>
<td>23.1</td>
<td>46.8</td>
<td>49.1</td>
<td>55.2</td>
<td>17.5</td>
<td>37.8</td>
<td>26.5</td>
<td>51.4</td>
<td>44.4</td>
<td>54.1</td>
<td>47.1</td>
<td>14.7</td>
<td>45.3</td>
<td>23.1</td>
<td>38.3</td>
<td>39.1</td>
<td>38.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(c) Validation 2007</th>
<th>Aero</th>
<th>Bike</th>
<th>Bird</th>
<th>Boat</th>
<th>Bottle</th>
<th>Bus</th>
<th>Car</th>
<th>Cat</th>
<th>Chair</th>
<th>Cow</th>
<th>Table</th>
<th>Dog</th>
<th>Horse</th>
<th>Mbike</th>
<th>Person</th>
<th>Plant</th>
<th>Sheep</th>
<th>Sofa</th>
<th>Train</th>
<th>TV</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCNN</td>
<td>81.1</td>
<td>80.1</td>
<td>70.2</td>
<td>53.7</td>
<td>43.0</td>
<td>71.2</td>
<td>71.3</td>
<td>80.1</td>
<td>61.5</td>
<td>81.3</td>
<td>62.7</td>
<td>81.1</td>
<td>81.6</td>
<td>80.5</td>
<td>50.7</td>
<td>33.8</td>
<td>70.8</td>
<td>72.7</td>
<td>81.5</td>
<td>72.6</td>
<td>69.1</td>
</tr>
<tr>
<td>GET.Loc</td>
<td>80.0</td>
<td>80.2</td>
<td>79.0</td>
<td>70.0</td>
<td>47.8</td>
<td>71.0</td>
<td>71.0</td>
<td>79.1</td>
<td>66.9</td>
<td>81.3</td>
<td>71.4</td>
<td>80.0</td>
<td>80.1</td>
<td>79.5</td>
<td>56.9</td>
<td>41.7</td>
<td>69.4</td>
<td>80.3</td>
<td>79.7</td>
<td>81.4</td>
<td>72.3</td>
</tr>
<tr>
<td>Trace.Loc</td>
<td>79.5</td>
<td>80.1</td>
<td>69.7</td>
<td>70.4</td>
<td>50.8</td>
<td>71.3</td>
<td>70.7</td>
<td>79.5</td>
<td>58.8</td>
<td>80.7</td>
<td>71.6</td>
<td>79.6</td>
<td>80.5</td>
<td>79.6</td>
<td>56.4</td>
<td>40.5</td>
<td>70.3</td>
<td>80.7</td>
<td>79.5</td>
<td>81.2</td>
<td>71.6</td>
</tr>
<tr>
<td>GT.Loc</td>
<td>79.6</td>
<td>80.5</td>
<td>79.5</td>
<td>70.2</td>
<td>48.6</td>
<td>71.1</td>
<td>71.0</td>
<td>79.2</td>
<td>60.0</td>
<td>80.7</td>
<td>71.4</td>
<td>80.1</td>
<td>88.3</td>
<td>87.6</td>
<td>56.5</td>
<td>40.9</td>
<td>70.0</td>
<td>80.3</td>
<td>79.2</td>
<td>81.2</td>
<td>72.8</td>
</tr>
</tbody>
</table>
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

- Introduction

- Literature Review
  - Object Proposal Generation
  - Image Representation
  - Object Localization
  - Object Recognition

- Object Localization by Optimizing Convolutional Neural Network Detection Score using Generic Edge Features
  - Proposed Method
  - Experimental Results

- Conclusion
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