CSC420: Tutorial on object detection

Bin Yang
17.11.2016
Outline

- What’s object detection? And its challenges?
- Face detection (demo)
- General object detection
- 3D object detection (demo)
Outline

• What’s object detection? And its challenges?
  • Face detection (demo)
  • General (multi-class) object detection
  • 3D object detection (demo)
Object detection

• Problem definition: Where? What?

• Challenges: Speed issue; large appearance variance.
Outline

• What’s object detection? And its challenges?
• Face detection (demo)
• General (multi-class) object detection
• 3D object detection (demo)
Face detection

- Binary classification: face / non-face
- Speed matters.

Cascaded Classifier

- A 1 feature classifier achieves 100% detection rate and about 50% false positive rate.
- A 5 feature classifier achieves 100% detection rate and 40% false positive rate (20% cumulative) — using data from previous stage.
- A 20 feature classifier achieves 100% detection rate with 10% false positive rate (2% cumulative)
VJ face detection results
Today’s image
Aggregate channel features (P. Dollar, 2009)

- To handle more difficult scenario, we need more robust features with better representational power.
Aggregate channel features (B. Yang, 2014)

- Train view-specific detectors for each face view.
CNN based face detector (H. Qin, 2016)
Real-time demos

Reference: http://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/
Qin_Joint_Training_of.CVPR_2016_paper.pdf
Outline

• What’s object detection? And its challenges?
• Face detection (demo)
• General (multi-class) object detection
• 3D object detection (demo)
General object detection

• No longer binary classification.

• We need very good image representation.

• CNN+ImageNet provide us with that.
1000-class image classification:
Top1 accuracy: 57.1%
Top5 accuracy: 80.2%
Region-CNN  (R. Girshick, 2013)

1. Input image
2. Extract region proposals (~2k)
3. Compute CNN features
4. Classify regions
Region-CNN  (R. Girshick, 2013)

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Architecture</th>
<th>Checkpoint</th>
<th>Top-1 Accuracy</th>
<th>Top-5 Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inception-ResNet-v2</td>
<td>Code</td>
<td>inception_resnet_v2_2016_08_30.tar.gz</td>
<td>80.4</td>
<td>95.3</td>
</tr>
<tr>
<td>Inception V3</td>
<td>Code</td>
<td>inception_v3_2016_08_28.tar.gz</td>
<td>78.0</td>
<td>93.9</td>
</tr>
<tr>
<td>ResNet 152</td>
<td>Code</td>
<td>resnet_v1_152_2016_08_28.tar.gz</td>
<td>76.8</td>
<td>93.2</td>
</tr>
<tr>
<td>ResNet V2 200</td>
<td>Code</td>
<td>TBA</td>
<td>79.9*</td>
<td>95.2*</td>
</tr>
</tbody>
</table>

(*) Results quoted in ResNet paper.
Outline

- What's object detection? And its challenges?
- Face detection (demo)
- General (multi-class) object detection
- 3D object detection (demo)
Video demo

Link: http://3dimage.ee.tsinghua.edu.cn/files/XiaozhiChen/mono3d/cvpr16_demo.webm
Questions?

• What’s object detection? And its challenges?
• Face detection (demo)
• General (multi-class) object detection
• 3D object detection (demo)