R-CNN: Region-based CNN

A framework for object detection with ConvNets

Input image

Extract region proposals

Compute CNN features

Classify regions

- aeroplane? no.
- person? yes.
- tvmonitor? no.

Any proposal method
(boxes or masks)

Any architecture,
not just AlexNet
(including GoogLeNet)

SoftMax, SVM

Ross Girshick, Jeff Donahue, Sergio Guadarrama, Trevor Darrell, Jitendra Malik — UC Berkeley
### R-CNN-ZF baseline on ILSVR14

<table>
<thead>
<tr>
<th></th>
<th>val2</th>
<th>test</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILSVRC13 + AN</td>
<td>29.7%</td>
<td></td>
</tr>
<tr>
<td>ILSVRC14 + AN</td>
<td>31.2%</td>
<td></td>
</tr>
<tr>
<td>ILSVRC14 + ZF</td>
<td>32.1%</td>
<td></td>
</tr>
<tr>
<td>ILSVRC14 + ZF + BBOX</td>
<td>33.4%</td>
<td></td>
</tr>
<tr>
<td>ILSVRC14 + ZF + BBOX</td>
<td>34.5%</td>
<td></td>
</tr>
</tbody>
</table>

Take away from other entries in ILSVRC14: huge gains come from improving the CNN module.