**Compare R-CNN to Fast R-CNN**

Qinghong Lin

2018/10

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **R-CNN** | **Fast R-CNN** | **Faster R-CNN** | **Mask R-CNN**(Faster R-CNN+FCN) |
| **Insight** | ·Apply high-capacity **CNNs** to bottom-up region proposals to localize and segment objects.  ·When labeled trainging data is scarce,supervised pre-training + domain-specific fine-tuning | ·Run the CNN **just once** per image and **share computation***(RoIPool)*.  ·Combine three models into **one network**  (*single-stage training* *using multi-task loss*). | ·Using a network***(RPN)*** to generate a good region proposals depended on features map. | ·Extends Faster R-CNN by **adding a branch** for predicting segmentation **masks** on each RoI.  ·**RoIAlign:** (U*sing bilinear interpolation*) Avoid any quantization of the RoI boundaries or bins. |
| **model** | IMG_256 | IMG_256 | IMG_256 | IMG_256 |
| **input** | Image | Images with region proposals | Image | Image |
| **output** | Bounding boxes + labels for each object | | | Bounding boxes + labels + binary mask |
| **process** | 1.**Selective search** creates RoI(~2k) from image.  2.**Warping** turn RoI into fixed-size CNN input.  3.**CNN** extract feature for every RoI.  4.**SVM** classify object + **bounding-box regression** reduce localization errors. | 1.**CNN** extract feature for entire image only one.  2.**Selective Search** creates RoI from feature map.  3.**RoI pooling layer** extract fixed-length feature vector from feature map*.*  4.Through FC layers, **softmax classifier** classify object + **bounding-box regressors**. | 1.**CNN** extract feature for image.  2.Create many **anchors** *(from feature)* as **RPN** input,output a good region(*by cls+reg).*  3.**RoI pooling layer** extract fixed-length feature from feature map*(crop by proposals).*  4.**softmax classifier** + **bounding-box regressors**. |  |
| **drawback** | ·Training is a **mult-stage pipeline** and **expensive** in space*(disk)* and time.  ·Testing is **slow** because **each RoI pass CNN without share computation.** | ·The **region proposer***(Selective Search)* is slow. | ·**RoIPool** coarse spatial quantization for feature map(*Not designed for* ***pixel-to-pixel alignment***). |  |