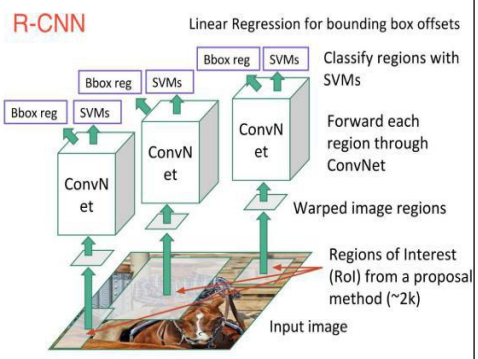
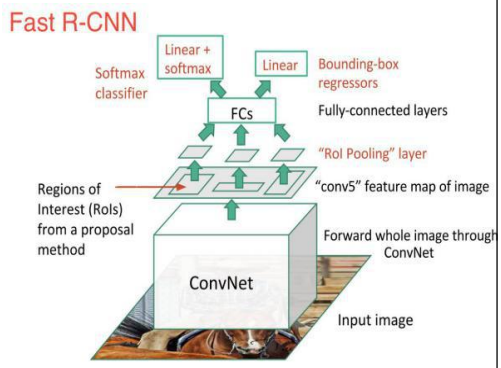
Compare R-CNN to Fast R-CNN

Qinghong Lin

2018/10

	R-CNN	Fast R-CNN
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*).
model		
input	Image	Images with region proposals
output	Bounding boxes + labels for each object	Tighter bounding boxes + labels for each object
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**.
result	**Train**: Fast R-CNN 9x faster than R-CNN*(VGG16)* **Testing**: Fast R-CNN 0.3s vs. 47s (per image) for R-CNN **mAP**: Fast R-CNN 66% vs.62% for R-CNN*(PASCAL VOC 2012)*	
drawback	1.Training is a **mult-stage pipeline** and **expensive** in space*(disk)* and time. 2.Testing is **slow** because **each RoI pass CNN without share computation.**	1.The **region proposer***(Selective Search)* is slow.