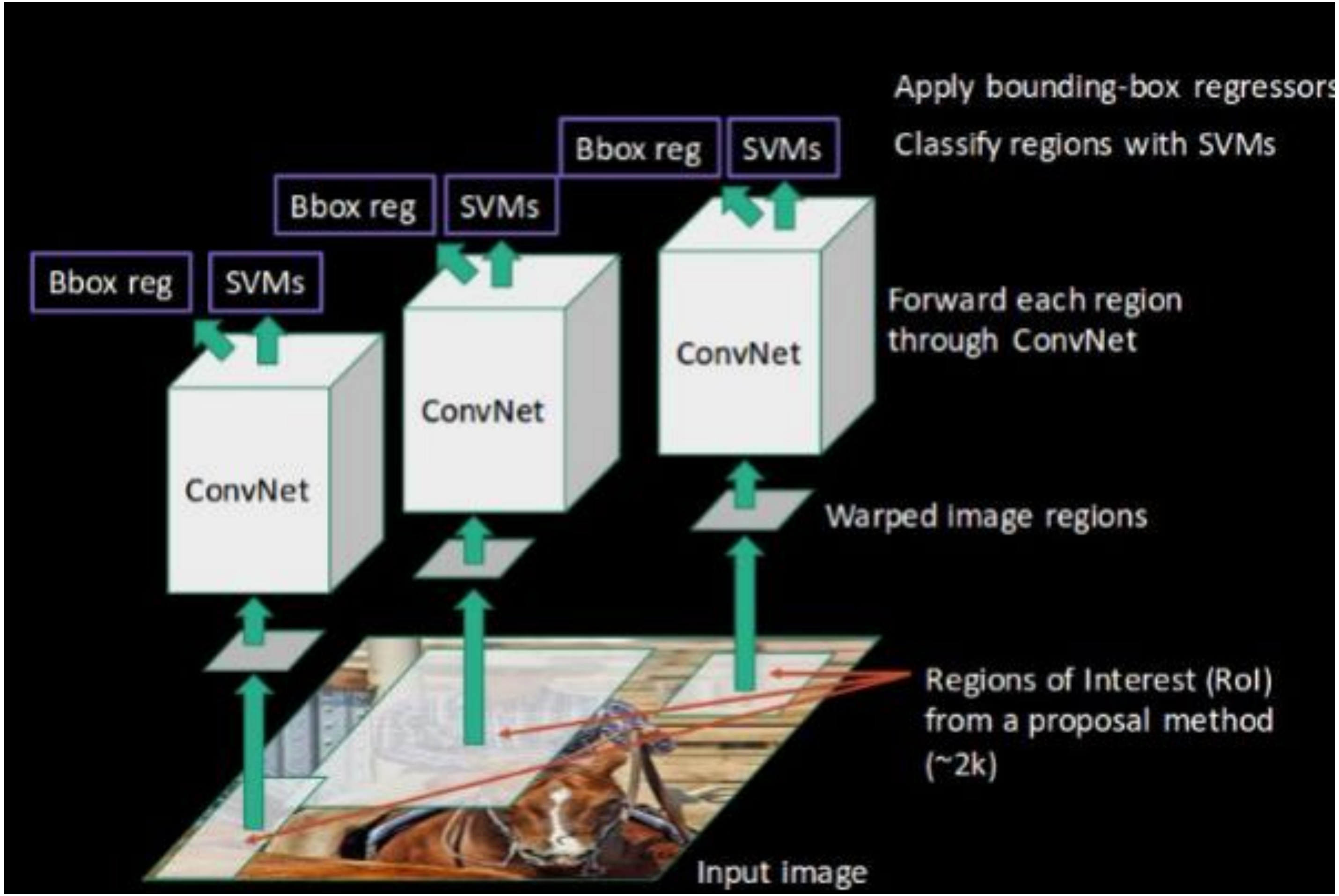
Object Detection

DNN-8.2 RCNN,Fast RCNN

Shahrukh khan

RCNN


Apply bounding-box regressors
Classify regions with SVMs
Bbox reg
SVMs
Bbox reg
SVMs
Bbox reg
SVMs
ConvNet
ConvNet
ConvNet
Forward each region through ConvNet
Warped image regions
Regions of Interest (RoI) from a proposal method (~2k)
Input image

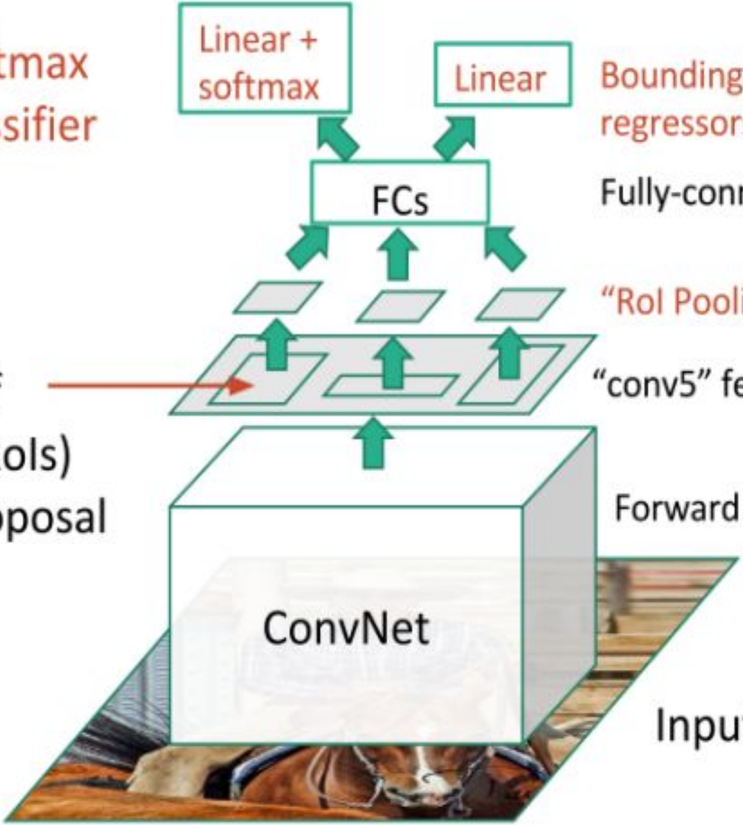
Limitation of RCNN

- Extracting 2,000 regions for each image based on selective search
- Extracting features using CNN for every image region. Suppose we have N images, then the number of CNN features will be N*2,000
- Inference (detection) is slow 47s / image with VGG16
- Training is multi stage pipeline.
- Training is expensive in space and time.

Fast R-CNN

- Training is single stage.
- No disk storage is required, end to end training.
- Improves training and testing speed.
- Increases detection accuracy
- 9x faster for training for VGG-16 than R-CNN
- 213x faster at test time than RCNN
- Implemented in C++ and caffee: https: //github.com/rbgirshick/fast-rcnn

Fast R-CNN



Softmax classifier

Bounding-box regressors

Fully-connected layers

"RoI Pooling" layer

"conv5" feature map of image

Regions of Interest (RoIs) from a proposal method

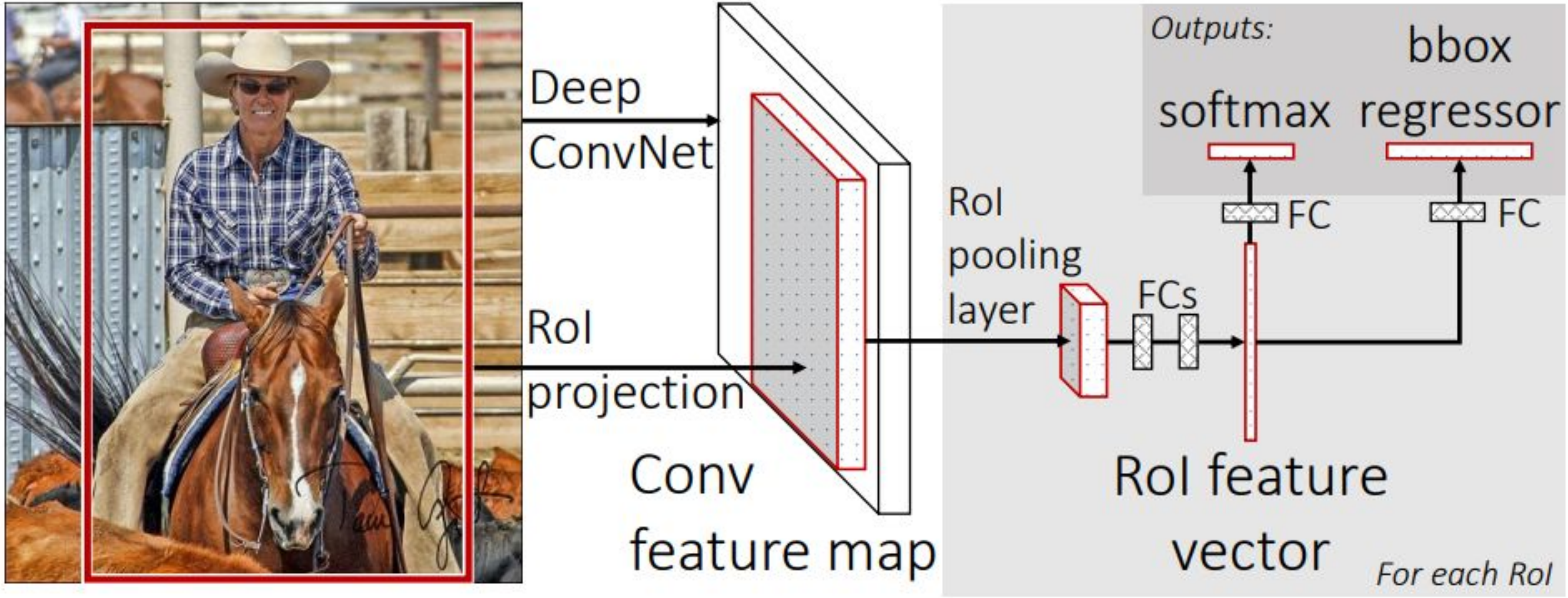
Forward whole image through ConvNet

Input image

Girshick, "Fast R-CNN", ICCV 2015.
Figure copyright Ross Girshick, 2015; source. Reproduced with permission.

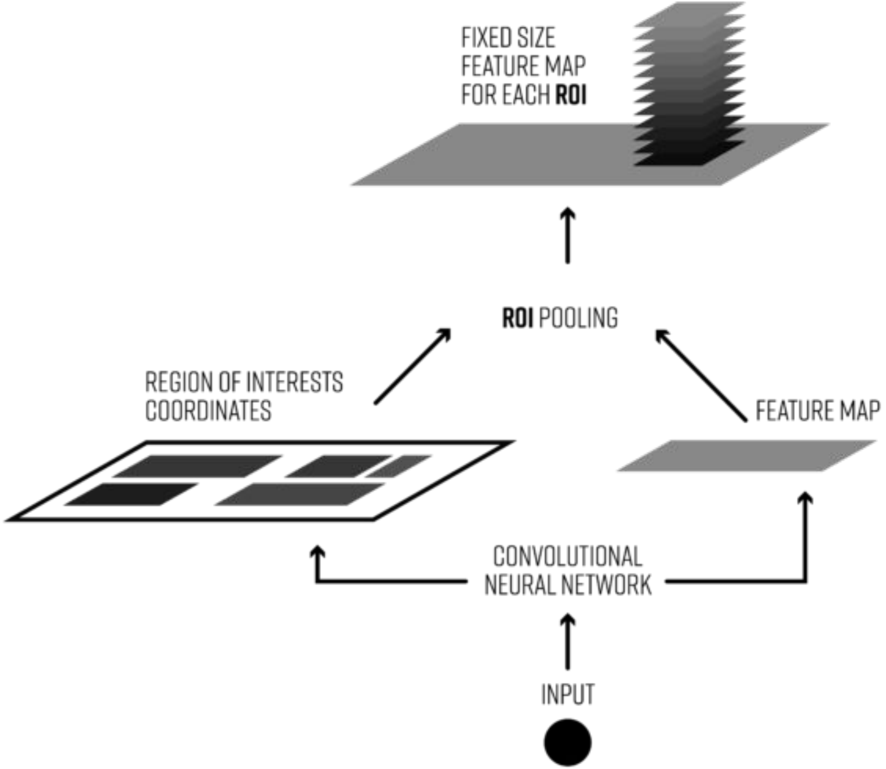
Fast RCNN : Architecture

- Takes input an entire image and set of object proposals.

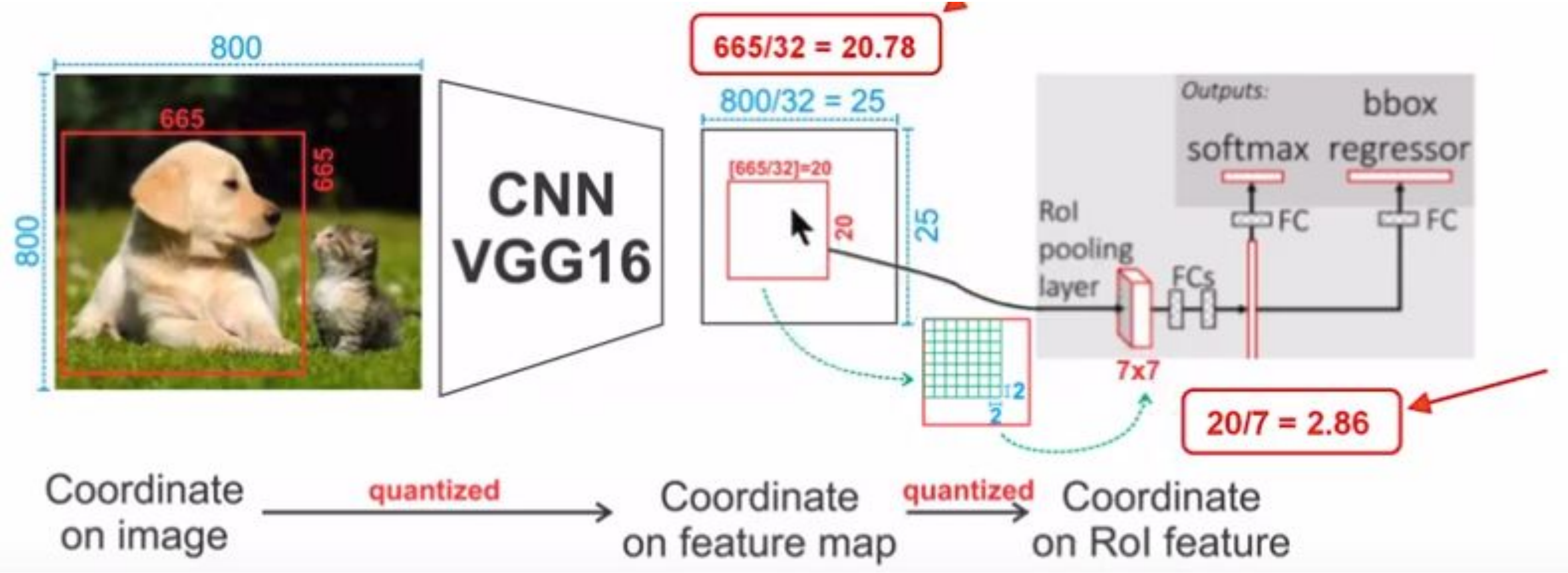


RoI Pooling layer

- ROI pooling layer uses max pooling to convert feature map into a fixed spatial extent H x W.
- Each RoI is a rectangular window into a conv feature map defined by (r, c, h, w).
- RoI max pooling works by dividing the h x w RoI into H x W grid of subwindows.
- Each subwindow of size (h / H x w / W).


FIXED SIZE FEATURE MAP FOR EACH ROI
ROI POOLING
REGION OF INTERESTS COORDINATES
FEATURE MAP
CONVOLUTIONAL NEURAL NETWORK
INPUT

RoI Pooling layer



Fast R-CNN : Architecture

- Last Pooling layer is replaced by ROI Pooling layer.
- Last fully connected is replaced by two sibling layers , for classification and regression.
- Network is modified to take two data inputs, images and ROI.

Fast R-CNN : Training

- SGD mini batches are sampled hierarchically,
  - First sample N images and then R/N Roi from each image.
  - Roi from same images share computation.
  - N = 2 , R = 128 i.e 64 ROI from each image.
  - 25% ROI with IOU > 0.5 i.e u>0
  - Remaining ROI are sampled from IOU in interval [0.1,0.5).
- Jointly optimize a softmax classifier over K+1 classes and bounding box regressor.
- In R-CNN , classifier , regressor and SVM are trained in separate stages.

Fast R-CNN : Training

- Truncated SVD
- For detection the number of RoIs to process is large and nearly half of the forward pass time is spent computing the fully connected layers.
- Large fully connected layers are easily accelerated by compressing them with truncated SVD
- Truncated SVD reduces the parameter count from uv to t(u+v).

Fast R-CNN : Multi Task Loss

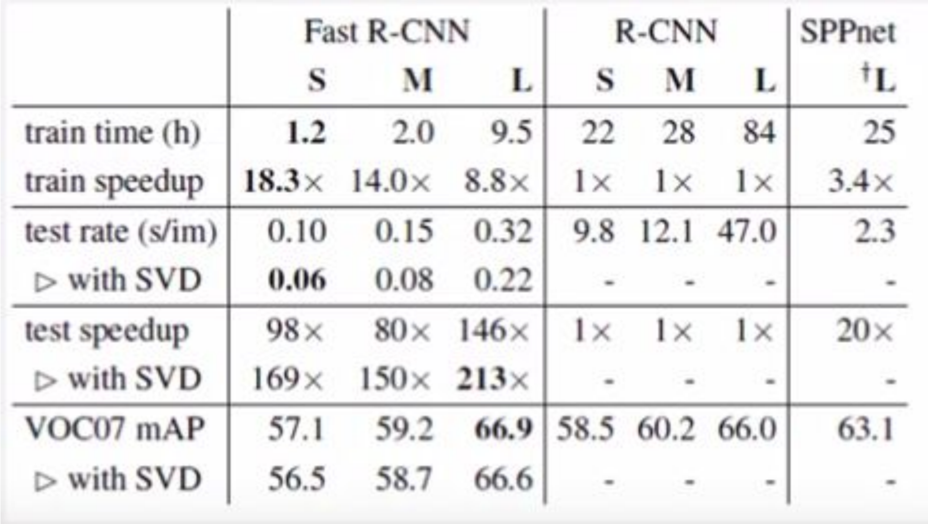
- u and v are ground truth for class and target bounding box.
- **L$_{\mathbf{cls}}$** and **L$_{\mathbf{loc}}$** are classification and regression loss.
- L$_{cls}$ = -*logp*$_{u}$,

$$\bullet\ L(p,u,t^u,v) = L_{cls}(p,u) + \lambda[u \geq 1]L_{loc}(t^u,v)$$

$$L_{\text{loc}}(t^u,v) = \sum_{i\in\{\text{x,y,w,h}\}} \text{smooth}_{L_1}(t_i^u - v_i),$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

Speed Comparison

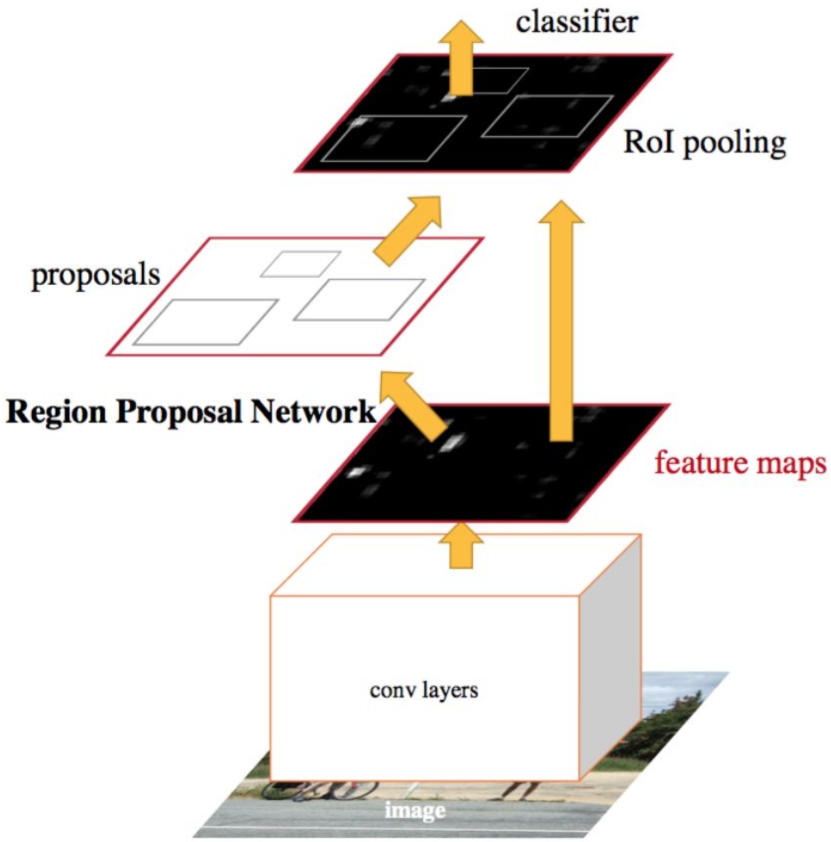


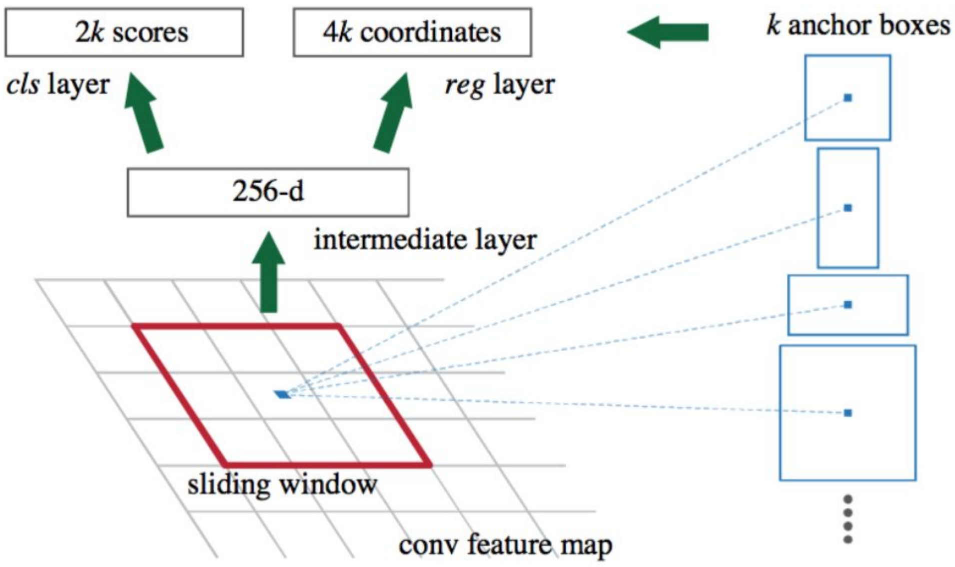
	Fast R-CNN			R-CNN			SPPnet
	S	**M**	**L**	**S**	**M**	**L**	†**L**
train time (h)	**1.2**	2.0	9.5	22	28	84	25
train speedup	**18.3**×	14.0×	8.8×	1×	1×	1×	3.4×
test rate (s/im)	0.10	0.15	0.32	9.8	12.1	47.0	2.3
▷ with SVD	**0.06**	0.08	0.22	-	-	-	-
test speedup	98×	80×	146×	1×	1×	1×	20×
▷ with SVD	169×	150×	**213**×	-	-	-	-
VOC07 mAP	57.1	59.2	**66.9**	58.5	60.2	66.0	63.1
▷ with SVD	56.5	58.7	66.6	-	-	-	-

Fast RCNN: Limitation

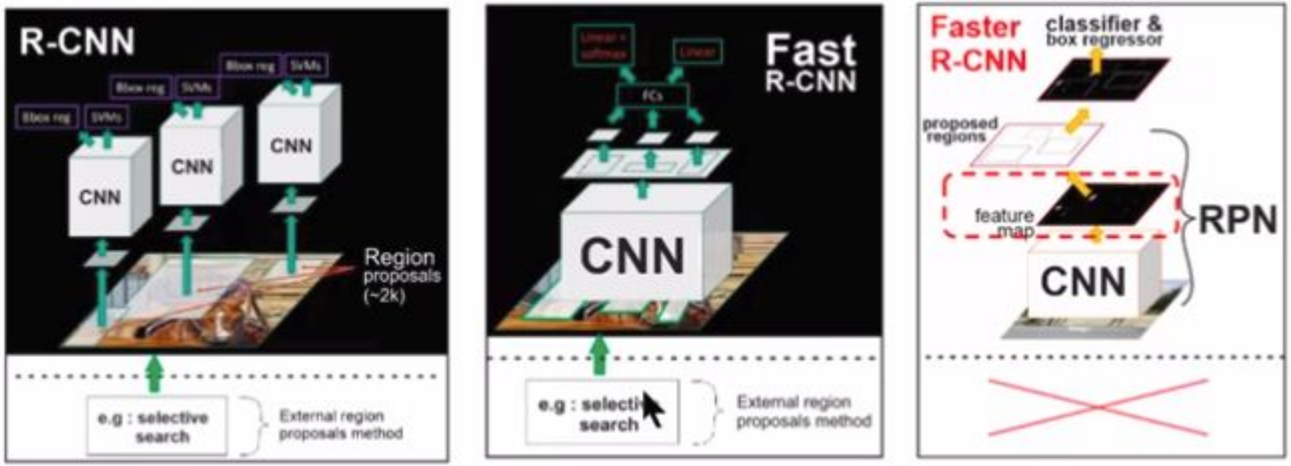
Region Proposals are still computational bottleneck. Selective search itself takes around 2 sec.

Faster RCNN


classifier
RoI pooling
proposals
Region Proposal Network
feature maps
conv layers
image


2*k* scores
4*k* coordinates
k anchor boxes
cls layer
reg layer
256-d
intermediate layer
sliding window
conv feature map

Comparison



	R-CNN	Fast R-CNN	Faster R-CNN
Test time per image	50 seconds	2 seconds	0.2 seconds
Speed-up	1x	25x	250x
mAP (VOC 2007)	66.0%	66.9%	66.9%