- Challenges in object detection :
    - Large number of regions to process.
    - Refining the boundaries of regions.
- SPP(Spatial pyramid pooling networks) were proposed to fasten RCNN by sharing computation for all regions of an image by using same feature map for all regions in that image.
- Features are extracted for a proposal by maxpooling the portion of the feature map inside the proposal into a fixed-size output (e.g., 6 × 6). Multiple output sizes are pooled and then concatenated as in spatial pyramid pooling.
- Even SPP is limited as there is no fine-tuning of previous layers which limits accuracy and still the architecture is multi-staged.
- In Fast RCNN, first feature map formed from image and then using RoI pooling layer for each region proposal, fixed length feature vector is obtained which is used to predict class probabilities which are characterised by softmax loss; and box co-ordinates use smoothL1 loss.
- In RoI pooling, a window of h*w is max-pooled to get fixed output grid of H*W.
- The RoI layer is simply the special-case of the spatial pyramid pooling layer used in SPPnets [11] in which there is only one pyramid level.
- Back-propagation through the SPP layer is highly inefficient when each training sample (i.e. RoI) comes from a different image, which is exactly how R-CNN and SPPnet networks are trained.
- In Fast RCNN, for a minibatch of N images, R/N RoIs are trained which share computation for each image.One problem is that the correlation between RoIs of same image may delay convergence.
- The regression targets are normalised to prevent gradient exploding due to unbound regression variables.
- Strategies for scale invariance:
    - Brute force : single scale.
    - Image pyramids: multi scale.
- SVD truncation was also used to reduce dimensionality of W to fasten detection.
- Fast RCNN achieves more accuracy than SPP net => fine-tuning improves acc more than multi-scale training.
- They hypothesised that fine-tuning only FCs would not help and their exp proved it by getting decrease in accuracy. So, they fine-tuned some of the convolutional layers that were task dependent (mainly later layers).
- Single scale processing had mAP close to multi-scale but offered great speed.
- mAP of Fast RCNN improves by 2-3% when dataset was increased => no high bias.
- Softmax outperformed SVM when used for classification in the multi-task loss.
- Sparse proposal detectors (selective search) act like cascade and removes most of the wrong proposals leaving small set for the network.It increases the mAP for both DPM and Fast RCNN.