Open Elective Course [OE]

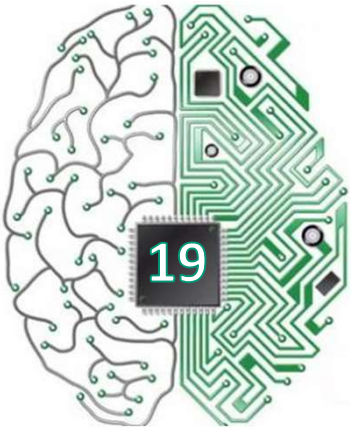
Course Code: CSO507

Winter 2023-24

Lecture#

Deep Learning

Unit-4: Convolutional Neural Networks (Part-VII)



Course Instructor:

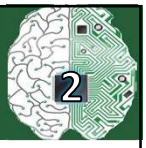
Dr. Monidipa Das

Assistant Professor

Department of Computer Science and Engineering

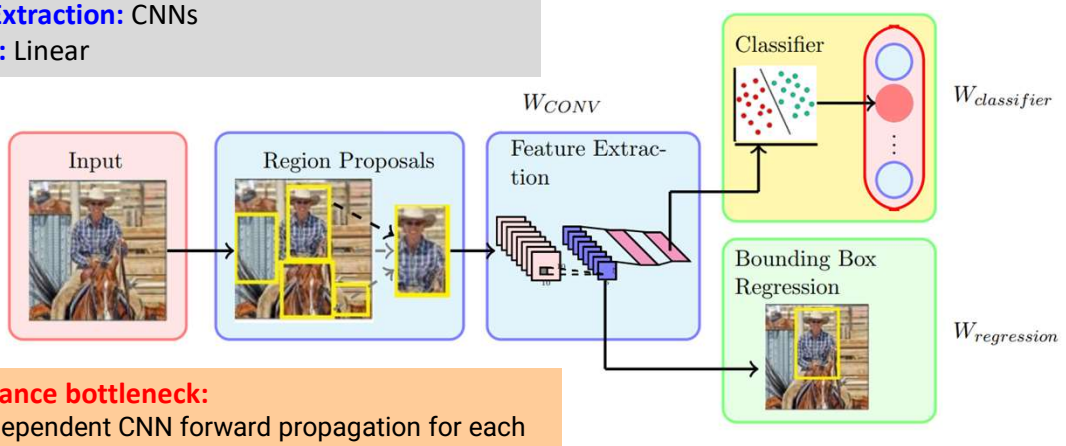
Indian Institute of Technology (Indian School of Mines) Dhanbad, Jharkhand 826004, India

RCNN model for object detection



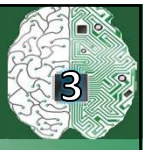
- **Region Proposals:** Selective Search (Uijlings *et al.*, 2013)
- **Feature Extraction:** CNNs
- **Classifier:** Linear

RCNN: (Girshick *et al.*, 2014)



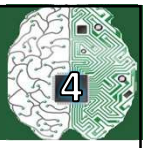
- **Performance bottleneck:**
  - Independent CNN forward propagation for each region proposal
  - No sharing computation.

Acknowledgement: Prof. Mitesh M. Khapra

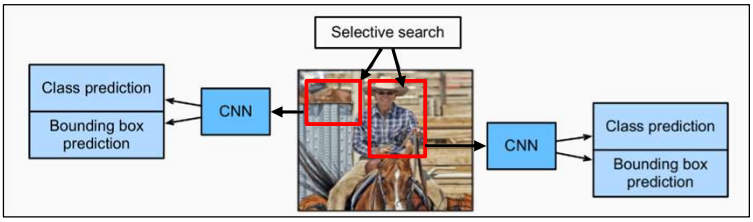


Fast RCNN model for object detection

RCNN vs. Fast RCNN



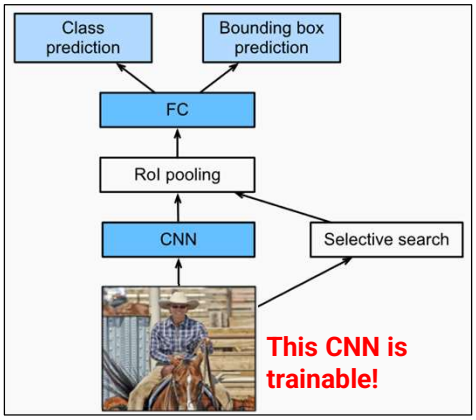
- **Performance improvement in Fast RCNN:**
  - CNN forward propagation is only performed on the entire image (Girshick, 2015).



The RCNN model

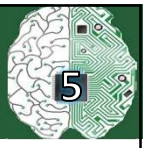
Object Detection Steps	RCNN	Fast RCNN
Region Proposals:	Selective Search	Selective Search
Feature Extraction:	CNNs	CNN
Classifier:	Linear	CNN

Source: https://d2l.ai/chapter_computer-vision/rcnn.html



The Fast RCNN model

Fast RCNN: Key Idea



- The fast R-CNN introduces the ***region of interest (RoI) pooling* layer**: the CNN output and region proposals are input into this layer, outputting concatenated features

  - Divide RoIs into k equally sized regions and do max pooling in each of those regions to construct a k dimensional vector
  - Connect the k dimensional vector to a fully connected layer
  - This max pooling operation is call RoI pooling

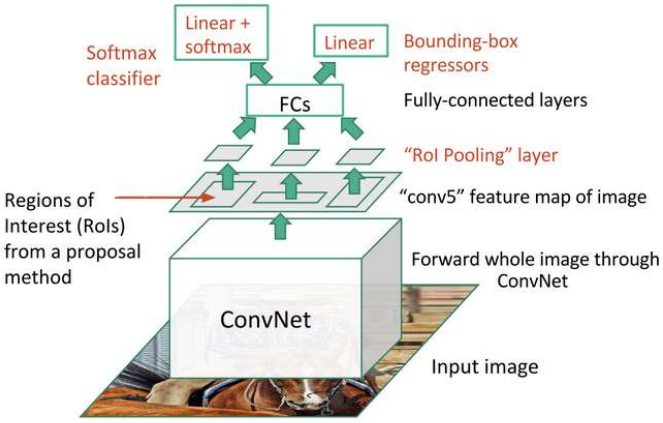
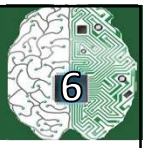


Image Source: Internet

Acknowledgement: Prof. Mitesh M. Khapra

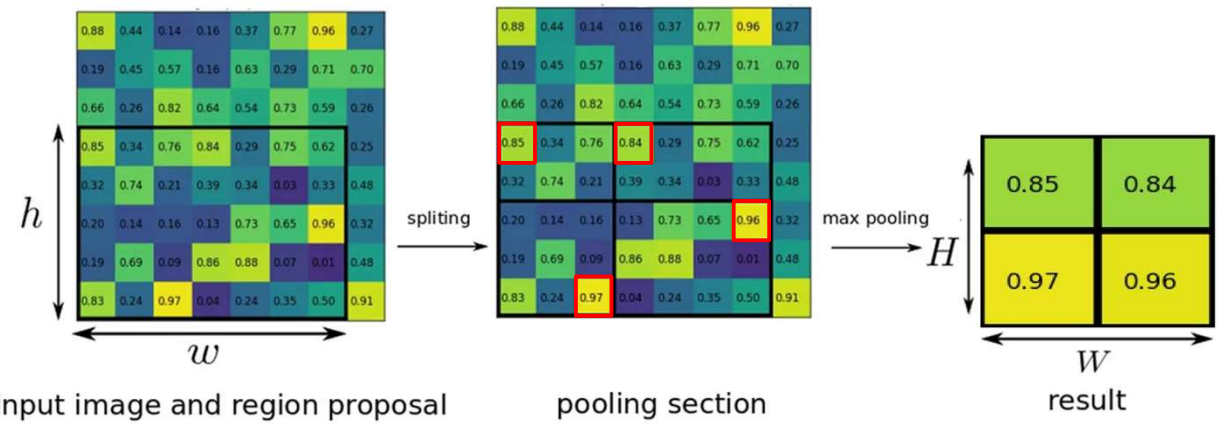
ROI Pooling



- Region of interest pooling (also known as RoI pooling) is an operation widely used in object detection tasks using convolutional neural networks.
- The operation was proposed in Fast RCNN paper in 2015.
- Its purpose is to perform max pooling on inputs of non-uniform sizes to obtain fixed-size feature maps.
- ROI pooling employes three steps to transform the input regions to similar size feature vectors:
  - Divide the region proposal into equal-sized sections (the number of which is the same as the dimension of the output).
  - Find the largest value in each section.
  - Copy these max values to the output buffer.

Acknowledgement: Ali Harakeh

ROI Pooling: Illustration

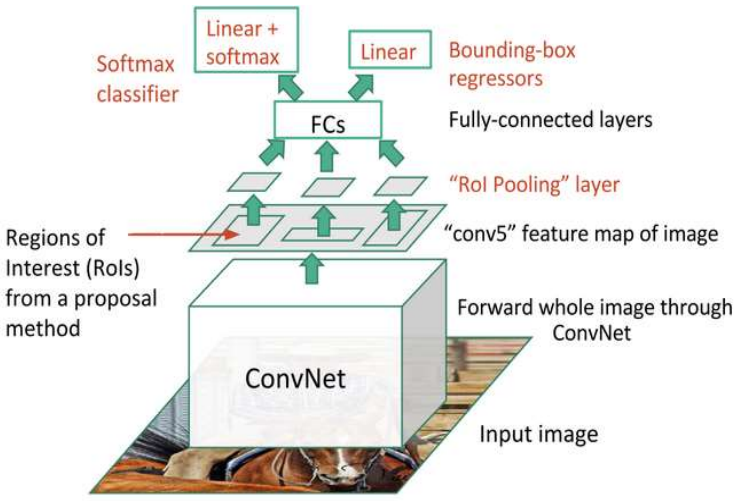


Acknowledgement: Ali Harakeh

Image Source: Internet

Fast RCNN

- Once we have the FC layer it gives us the representation of this region proposal
- We can then add a softmax layer on top of it to compute a probability distribution over the possible object classes
- Similarly we can add a regression layer on top of it to predict the new bounding box ($w*$, $h*$, $x*$, $y*$)



Acknowledgement: Prof. Mitesh M. Khapra

Image Source: Internet

Fast RCNN → Faster RCNN

- **Fast RCNN**
  - **Region Proposals:** Selective Search
  - **Feature Extraction:** CNN
  - **Classifier:** CNN

Faster RCNN:-
Idea: ***Can CNN be used for making region proposals also?***

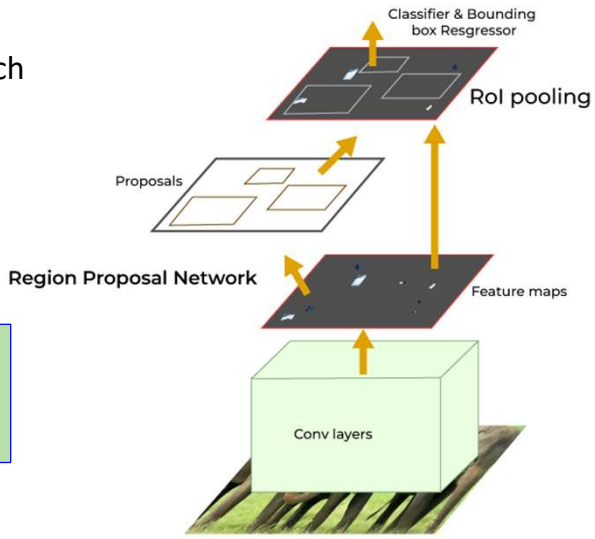
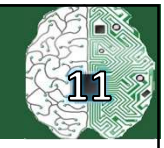


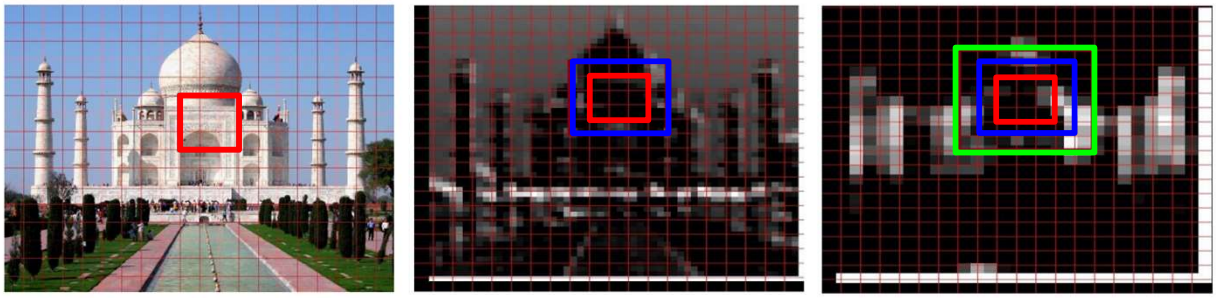
Image Source: Internet

Faster RCNN model for object detection

Faster RCNN: Key Idea

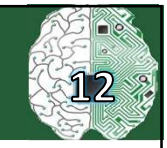


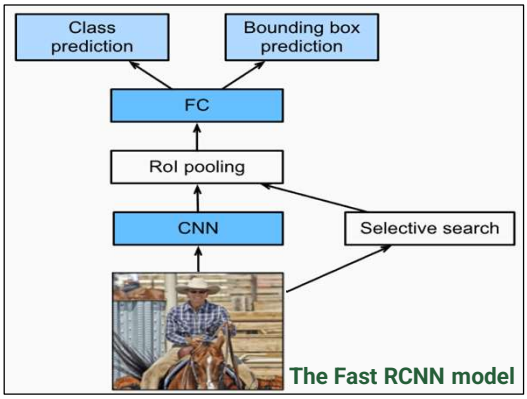
- We can get a **bounding box's region of influence on any layer in the CNN**

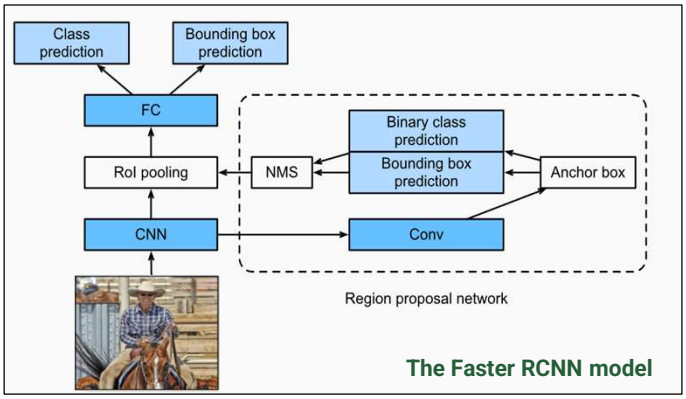


Acknowledgement: Prof. Mitesh M. Khapra

Fast RCNN vs. Faster RCNN





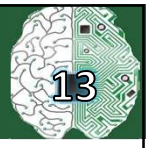


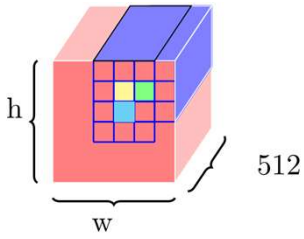
- **Performance improvement in Faster RCNN:**
  - Selective search is replaced by a region proposal network (RPN) [Ren et al., 2015].
  - Reduce region proposals without loss of accuracy

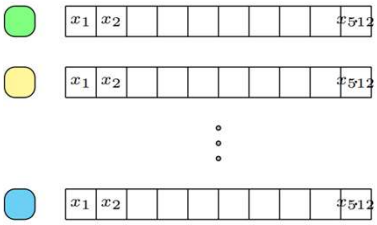
Source: https://d2l.ai/chapter_computer-vision/rcnn.html

Object Detection Steps	Fast RCNN	Faster RCNN
Region Proposals:	Selective Search	CNN
Feature Extraction:	CNN	CNN
Classifier:	CNN	CNN

Faster RCNN: Region Proposal

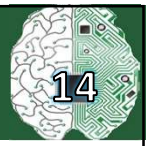


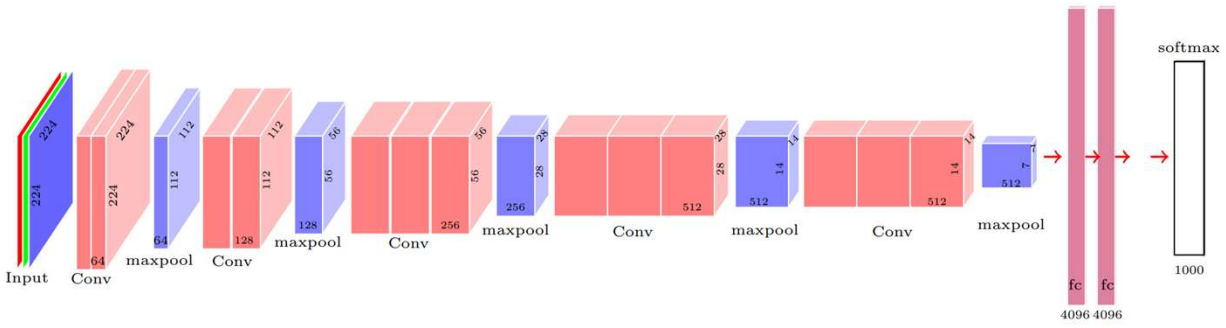




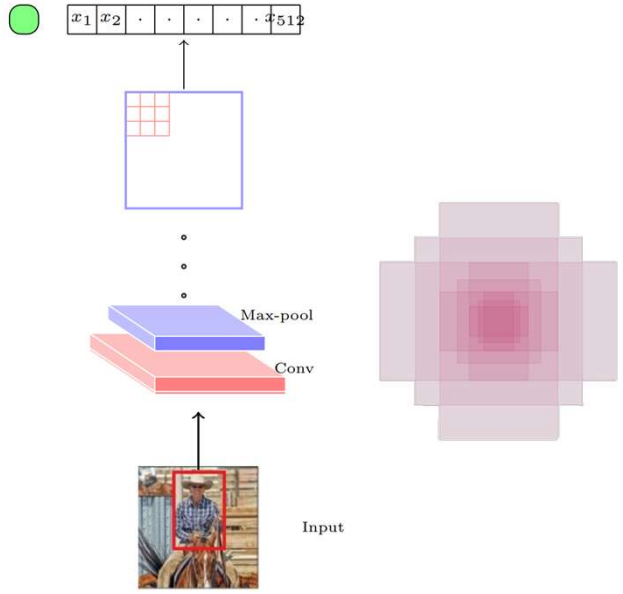
- Consider the output of the last convolutional layer of VGGNet
- Now consider one cell in one of the 512 feature maps
- If we apply a 3 ×3 kernel around this cell then we will get a 1D representation for this cell
- If we repeat this for all the 512 feature maps then we will get a 512 dimensional representation for this position
- We use this process to get a 512 dimensional representation for each of the w × h positions

VGGNet



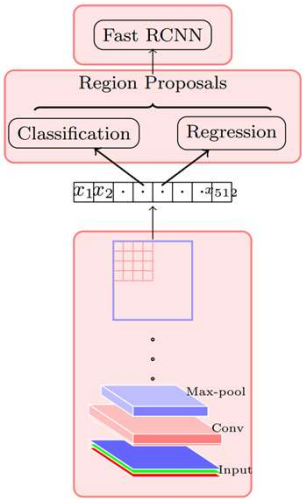


Faster RCNN: Region Proposal



- Consider k bounding boxes (called anchor boxes) of different sizes & aspect ratio
- Need to answer the following two questions:
- Given the 512d representation of a position, what is the probability that a given anchor box centered at this position contains an object? (Classification)
- How do you predict the true bounding box from this anchor box? (Regression)

Faster RCNN



Faster RCNNTraining:

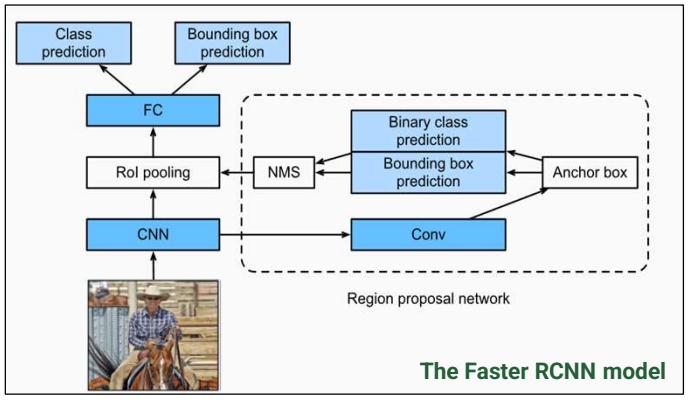
- Fine-tune RPN using a pre-trained ImageNet network
- Fine-tune fast RCNN from a pretrained ImageNet network using bounding boxes from step 1
- Keeping common convolutional layer parameters fixed from step 2, finetune RPN (post conv5 layers)
- Keeping common convolution layer parameters fixed from step 3, finetune fc layers of fast RCNN

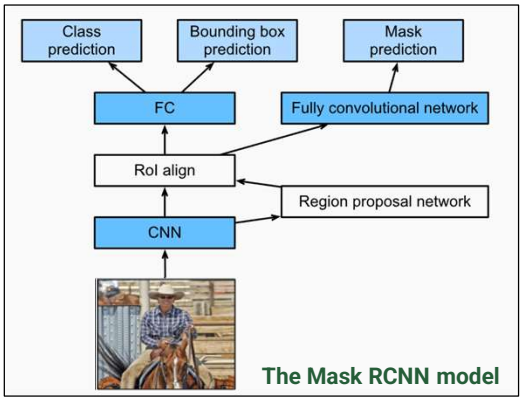
17

Mask RCNN

Faster RCNN vs. Mask RCNN

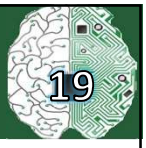
18



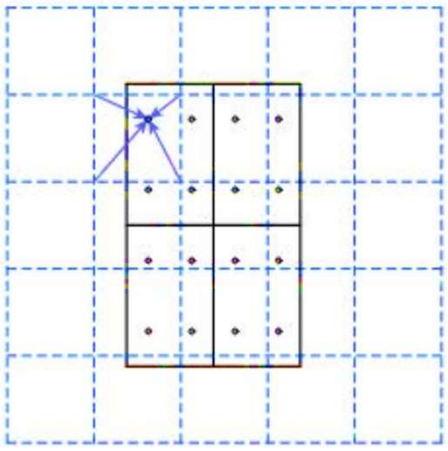


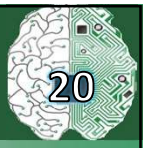
- Mask R-CNN replaces the region of interest pooling layer with the ***region of interest (RoI) alignment*** layer.

RoIAlign



- To extract the pixel-pixel mask, the RoI to be well aligned to preserve the explicit per-pixel spatial correspondence
- **RoIPool:** quantize a floating-number RoI to the discrete granularity of the feature map
- **RoIAlign:** bilinear interpolation to compute the exact values of the input features





Questions?