Open Elective Course [OE]

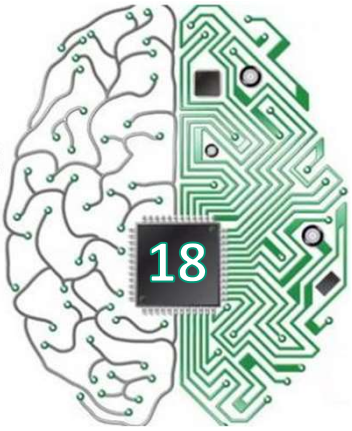
Course Code: CSO507

Winter 2023-24

Lecture#

Deep Learning

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



Course Instructor:

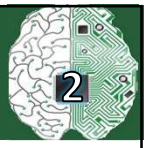
Dr. Monidipa Das

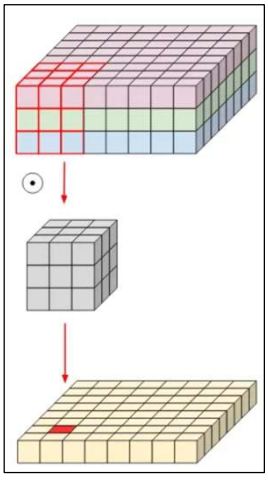
Assistant Professor

Department of Computer Science and Engineering

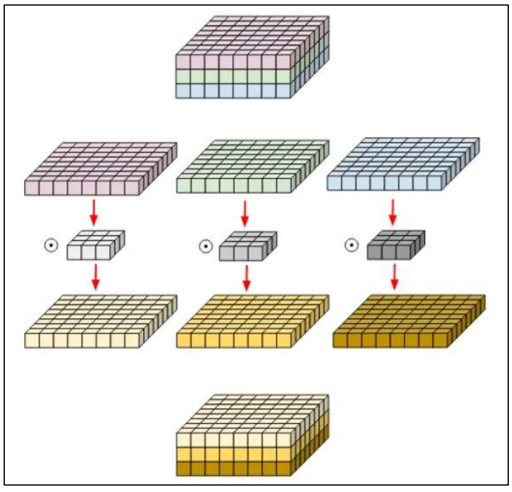
Indian Institute of Technology (Indian School of Mines) Dhanbad, Jharkhand 826004, India

Illustrations for Standard, Depth-wise, and Depth-wise Separable Convolution

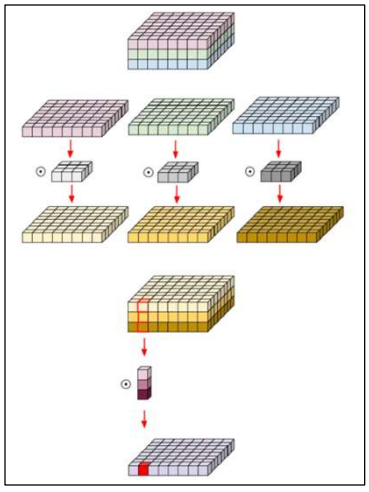




Standard convolution



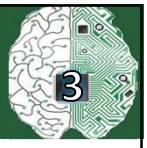
Depth-wise convolution



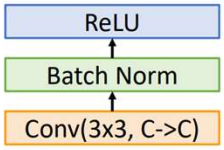
Depth-wise separable convolution

Source: Internet

MobileNets: Tiny Networks (For Mobile Devices)



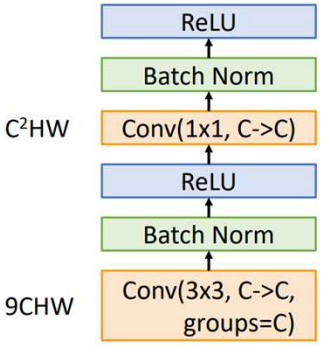
Standard Convolution Block
Total cost: $9C^2HW$



$9C^2HW$

Speedup = $9C^2/(9C+C^2)$
= $9C/(9+C)$
=> 9 (as C->inf)

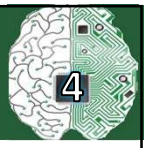
Depthwise Separable Convolution
Total cost: $(9C + C^2)HW$



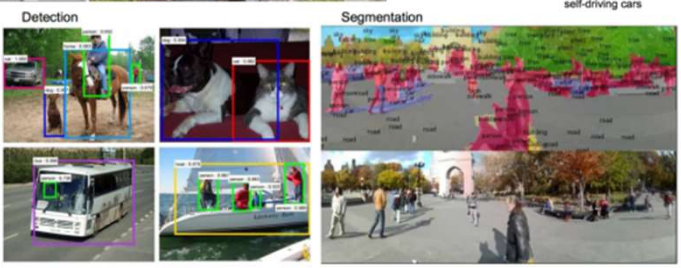
C^2HW "Pointwise Convolution"

9CHW "Depthwise Convolution"

ConvNets Application



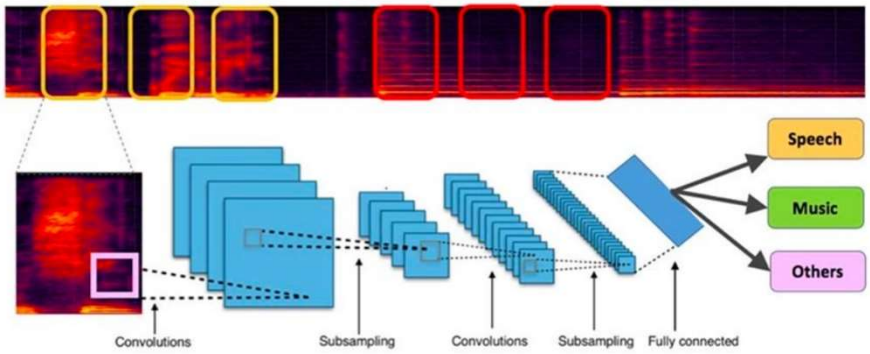
- ConvNets are today ubiquitous in **computer vision**!



ConvNets beyond vision

5

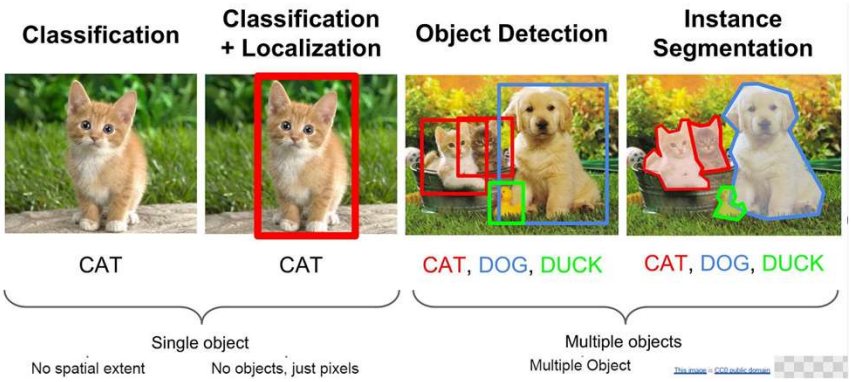
- CNNs are not only useful for image tasks!
- They are becoming the standard in audio tasks and very competitive in text processing tasks (e.g., sentiment classification).



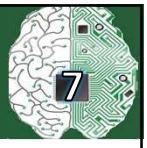
ConvNets Application for Computer Vision

6

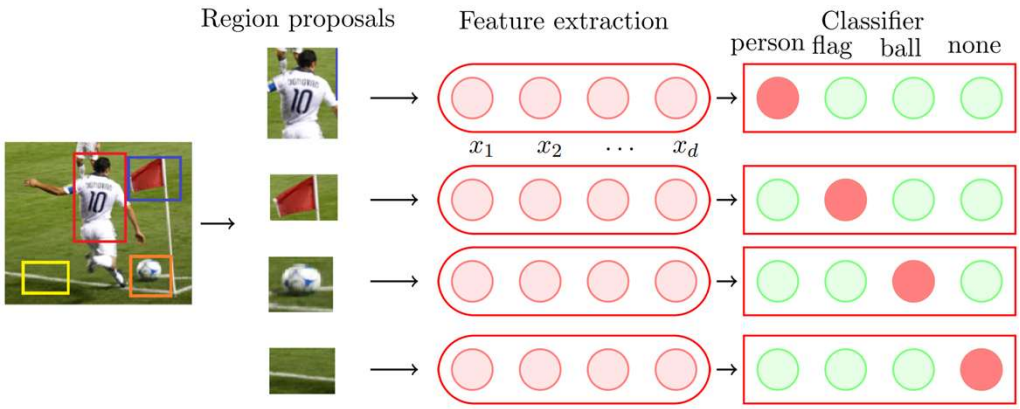
- Four major computer vision problems:
  - Image Classification
  - Image Classification with Localization
  - Object Detection
  - Object Segmentation



Pipeline for Object Detection

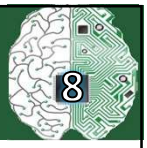


- Starts with a region proposal stage where we identify potential regions which may contain objects

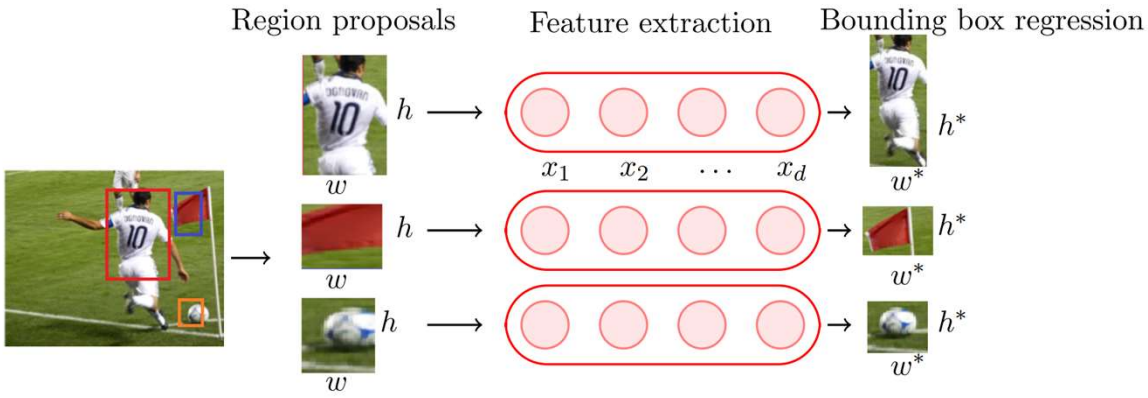


Acknowledgement: Prof. Mitesh M. Khapra

Pipeline for Object Detection



- In addition we would also like to correct the proposed bounding boxes This is posed as a regression problem (for example, we would like to predict w $*$, h $*$ from the proposed w and h)



Acknowledgement: Prof. Mitesh M. Khapra

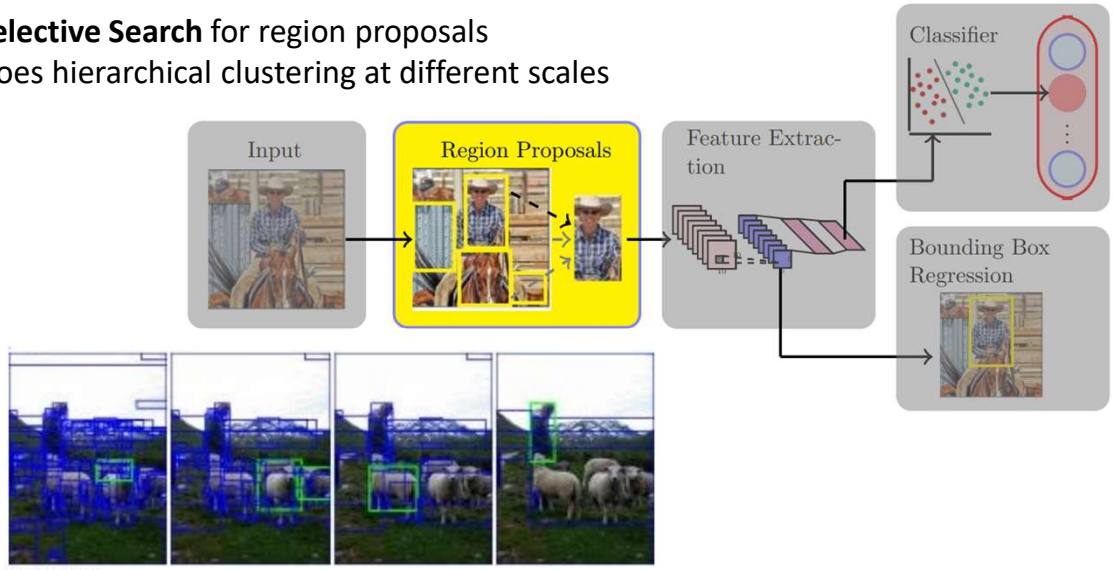
9

Region-based CNNs (RCNNs)

10

RCNN model for object detection

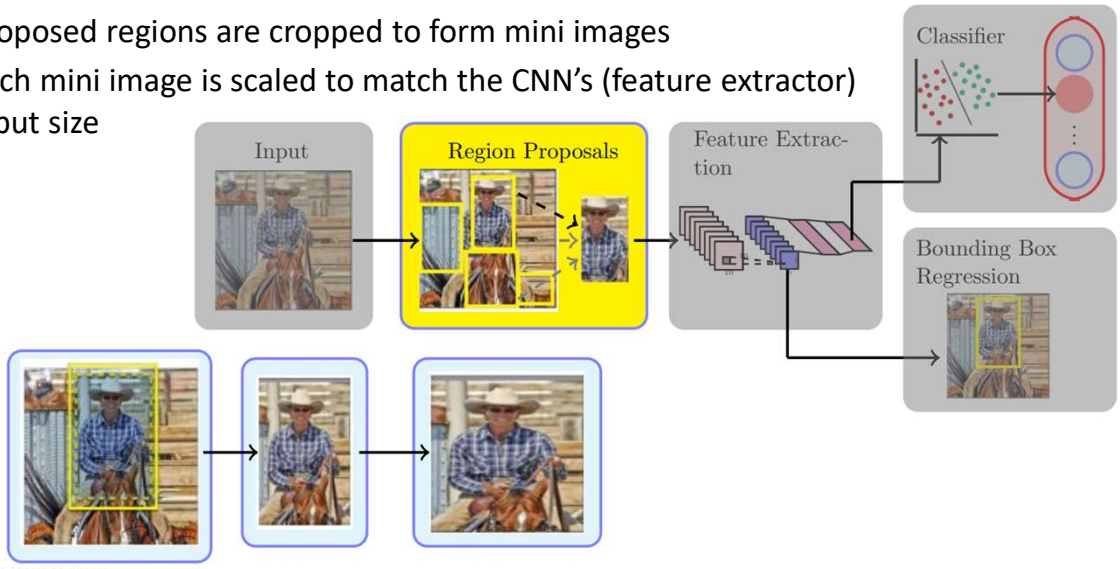
- **Selective Search** for region proposals
- Does hierarchical clustering at different scales



RCNN model for object detection

11

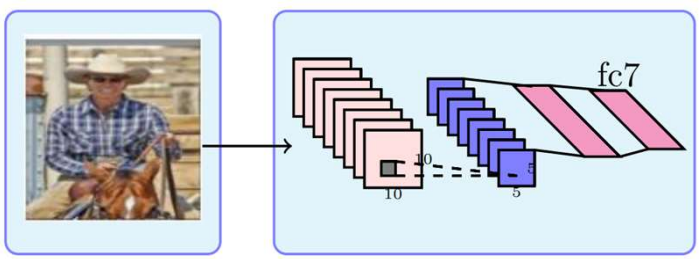
- Proposed regions are cropped to form mini images
- Each mini image is scaled to match the CNN's (feature extractor) input size



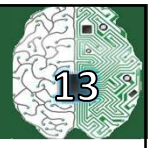
RCNN model for object detection

12

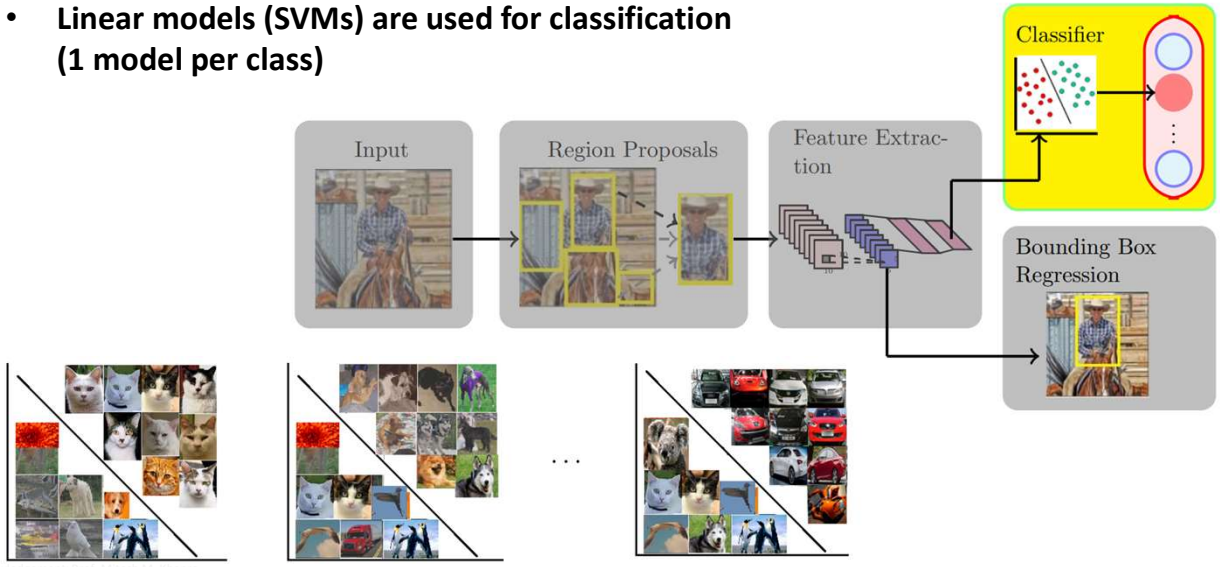
- For **feature extraction** any CNN trained for Image Classification can be used (AlexNet/ VGGNet etc.)
- Outputs from fc7 layer are taken as features
- CNN is fine tuned using ground truth (cropped) object images



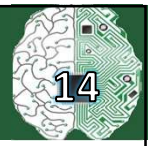
RCNN model for object detection



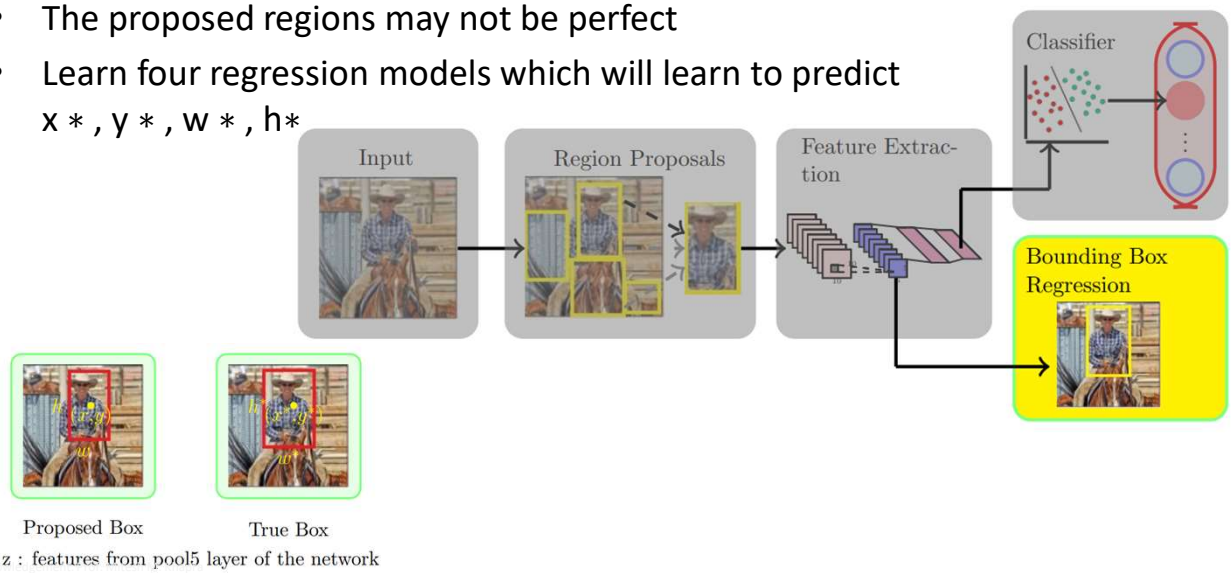
- **Linear models (SVMs) are used for classification (1 model per class)**



RCNN model for object detection

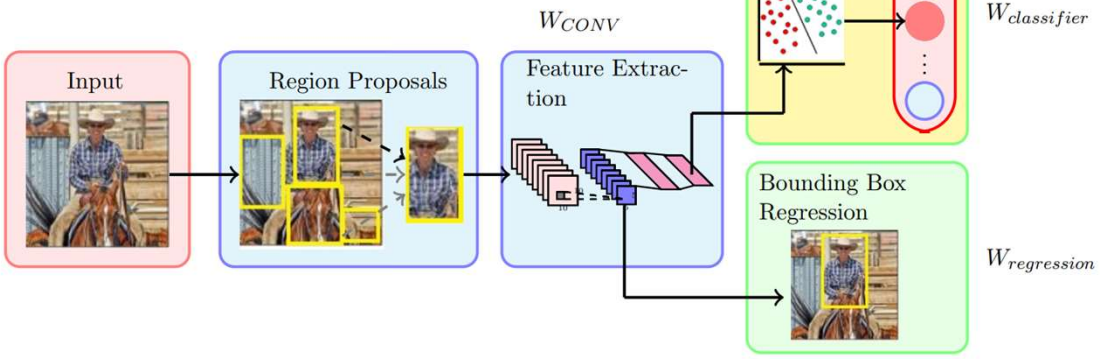


- The proposed regions may not be perfect
- Learn four regression models which will learn to predict $x*, y*, w*, h*$



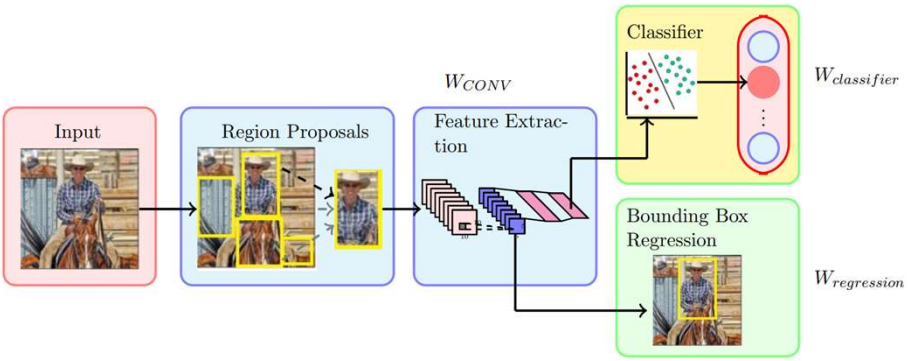
RCNN model for object detection

- What are the parameters of this model?
- W_{CONV} is taken as it is from a CNN trained for Image classification (say on ImageNet)
- W_{CONV} is then fine tuned using ground truth (cropped) object images
- $W_{classifier}$ is learned using ground truth (cropped) object images
- $W_{regression}$ is learned using ground truth bounding boxes

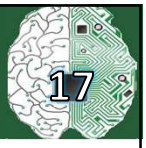


RCNN model for object detection

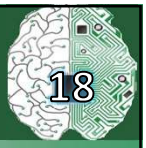
- **What is the computational cost for processing one image at test time?**
  - **Inference Time =** Proposal Time + # Proposals × Convolution Time + # Proposals × classification + # Proposals × regression



RCNN model for object detection



- No joint learning
- Use ad hoc training objectives
  - Fine tune network with softmax classifier (log loss)
  - Train post-hoc linear SVMs (hinge loss)
  - Train post-hoc bounding-box regressors (squared loss)
- Training (≈ 3 days) and testing (47s per image) is slow
- Takes a lot of disk space



Questions?