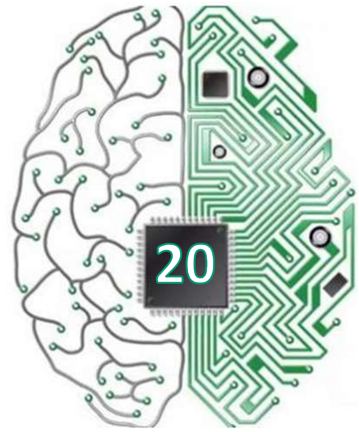
Open Elective Course [OE]
Course Code: CSO507
Winter 2023-24

Lecture#



Deep Learning

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

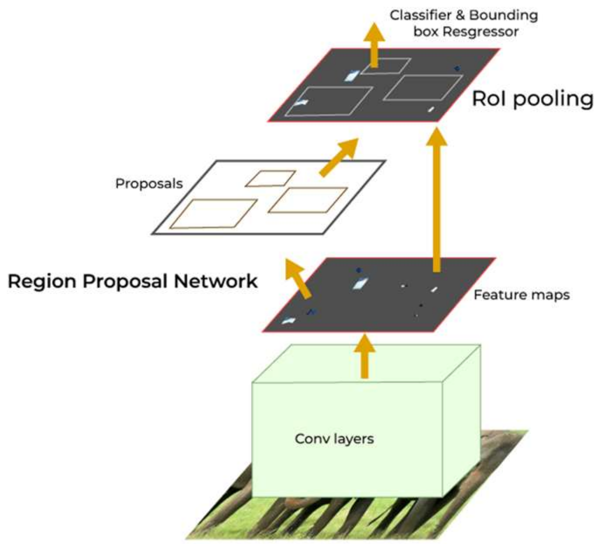
Unit-5: Sequence Modeling

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

2

YOLO Model for Object Detection

YOLO: You Only Look Once

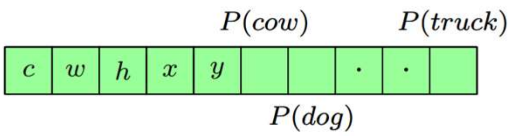


- The approaches that we have seen so far are two stage approaches
- They involve a region proposal stage followed by separate classification and regression stage
- Can we have an end-to-end architecture which does both proposal and classification simultaneously ?
- This is the idea behind YOLO
  - **You Only Look Once.**

Key Insights

- Previous Approaches
  - more complicated model pipeline
  - expensive computation
  - lacks contextual information for detection
- YOLO Algorithm
  - less complicated pipeline
  - efficient computation
  - has contextual information for detection

YOLO: You Only Look Once

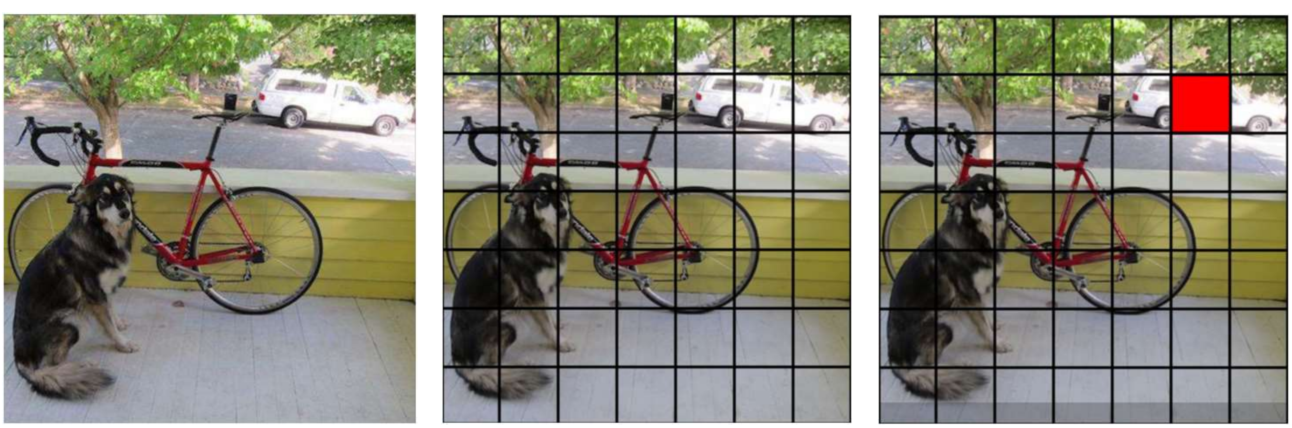



S × S grid on input

Considering only 1 bounding box per cell

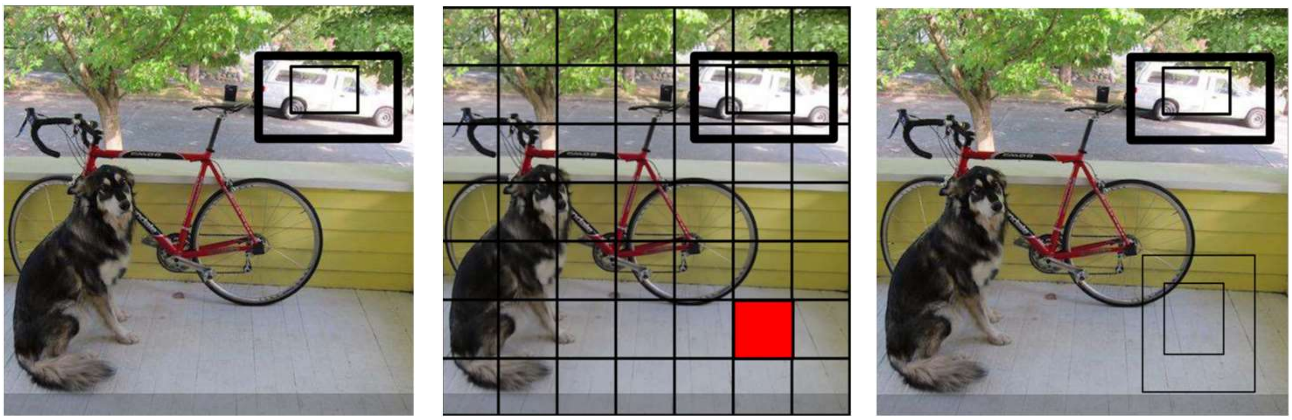
- Divide an image into S × S grids
- For each such cell we are interested in predicting 5 + k quantities
  - Confidence
  - Width of the bounding box
  - Height of the bounding box
  - Center (x,y) of the bounding box
  - Probability of the object in the bounding box belonging to the k-th class (k - values)
- The output layer thus contains S × S × (5 + k) elements

YOLO: You Only Look Once



YOLO: You Only Look Once

7



YOLO: You Only Look Once

8

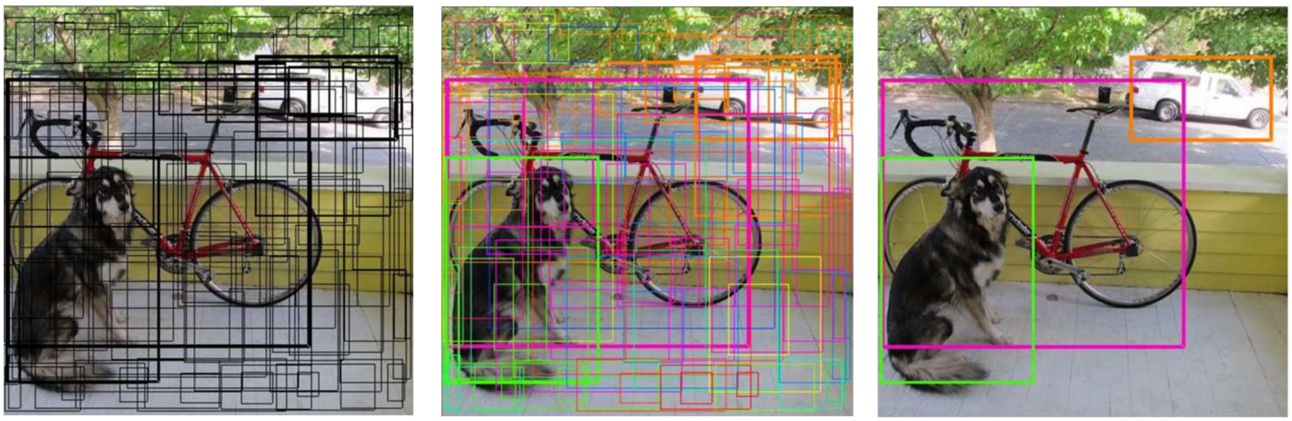
Considering B number of bounding boxes per cell, the output layer dimension would be: $\boldsymbol{S \times S \times (B * 5 + k)}$

YOLO: You Only Look Once

9

Class-specific Confidence

After applying Non-Maximum Suppression (NMS)

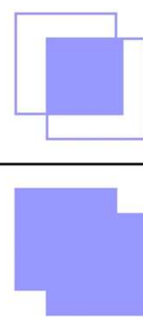


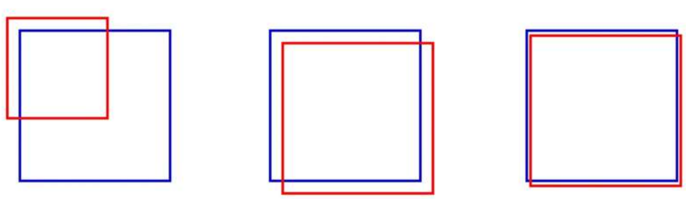
Class-specific Confidence: $\Pr(\text{Class}_i|\text{Object}) * \Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} = \Pr(\text{Class}_i) * \text{IOU}_{\text{pred}}^{\text{truth}}$

IoU: Intersection over Union

10

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

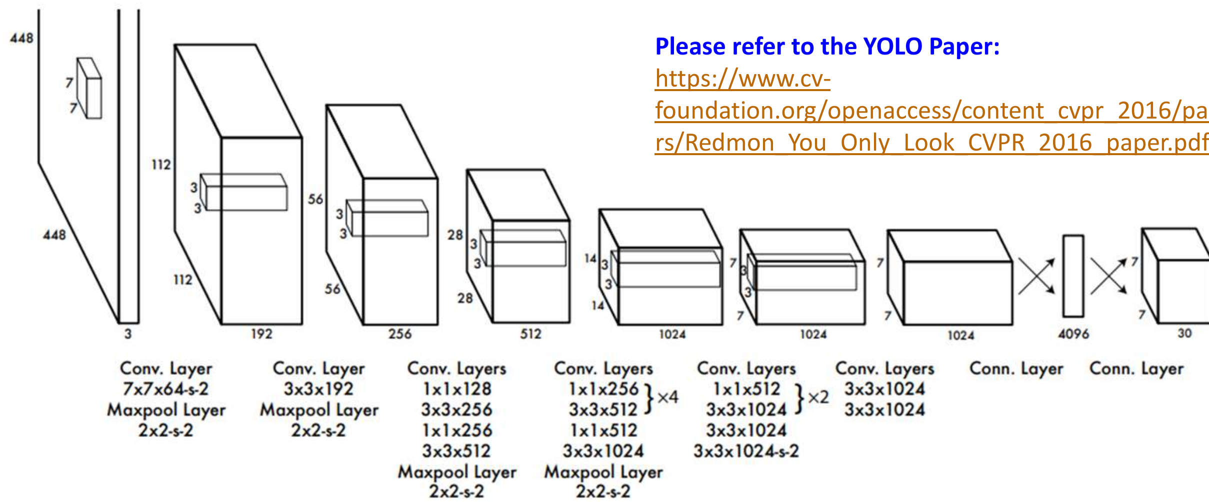




Poor Good Excellent

Network Design

Please refer to the YOLO Paper:
https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf



Inspired from GoogLeNet, however does not use Inception layer

YOLO Objective Function

$$
\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]
$$

$$
+ \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]
$$

$$
+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2
$$

$$
+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2
$$

$$
+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c) \right)^2
$$

Please refer to the YOLO Paper:
https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf

YOLO Objective Function

13

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$
$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

Coordinate Loss: Minimize the difference between x,y,w,h pred and x,y,w,h ground truth. ONLY IF object exists in grid box and if bounding box is resp for pred

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

Confidence Loss: Loss based on confidence ONLY IF there is object

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

No Object Loss: based on confidence if there is no object

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

Class loss, minimize loss between true class of object in grid box

Advantages of YOLO

14

- Pipeline comprised of a single network
- Learns general representation of the objects
- Extremely fast
- Reasons globally

Drawbacks of YOLO

15

- More localization error
- Loss function

16

Unit-5: Introduction (Sequence Modelling)

Sequential Data Examples

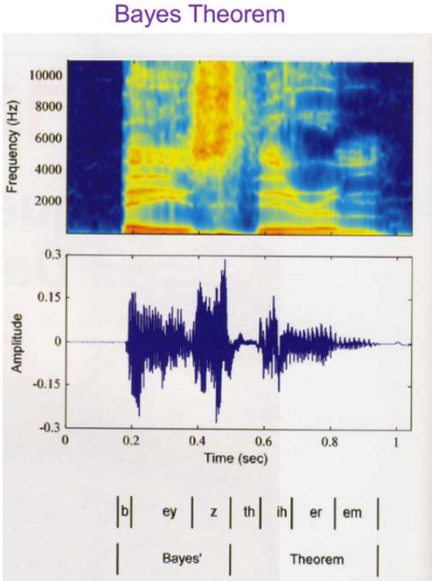
17

- Often arise through measurement of time series
  - Acoustic features at successive time frames in speech recognition
  - Sequence of characters in an English sentence
  - Parts of speech of successive words
  - Snowfall measurements on successive days
  - Rainfall measurements on successive days
  - Daily values of currency exchange rate
  - Nucleotide base pairs in a strand of DNA

Example

18

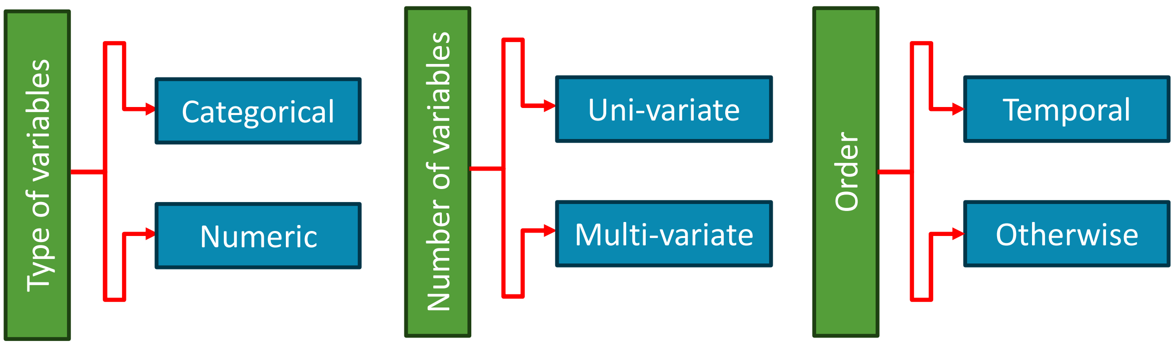
Bayes Theorem



- Decompose sound waves into frequency, amplitude using Fourier transforms
- Plot of the intensity of the spectral coefficients versus time index

Types of Sequential Data

19



Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

Two common tasks with sequential data

20

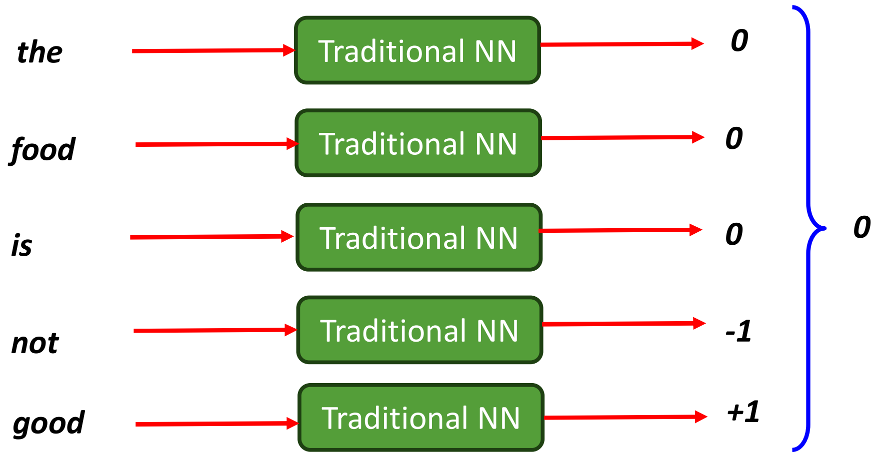
- **1. Sequence-to-sequence**
  - Named Entity Recognition
    - Input: Jim bought 300 shares of Acme Corp. in 2006
    - NER: [Jim]Person bought 300 shares of [Acme Corp.]Organization in [2006]Time
  - Machine Translation: Echte dicke kiste → Awesome sauce
- **2. Sequence-to-symbol**
  - Sentiment:
    - Best movie ever → Positive
  - Speaker recognition
    - Sound spectrogram → Harry

Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

Traditional NN & Sequence Data

21

- Categorizing a piece of text: *“**the food is not good**”*



22

Questions?

Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad