CSE4261: Neural Network and Deep Learning

Lecture: 29.06.2025

Sangeeta Biswas, Ph.D.
Associate Professor,
University of Rajshahi, Rajshahi-6205, Bangladesh

Training in Ultralytics

```python
# Load a pretrained model
model = YOLO("yolo11n.pt")
# Train the pretrained model
results = model.train(
        data = "train_val_test.yaml",
        epochs = 100,
        imgsz = 640,
        freeze = 10)
```

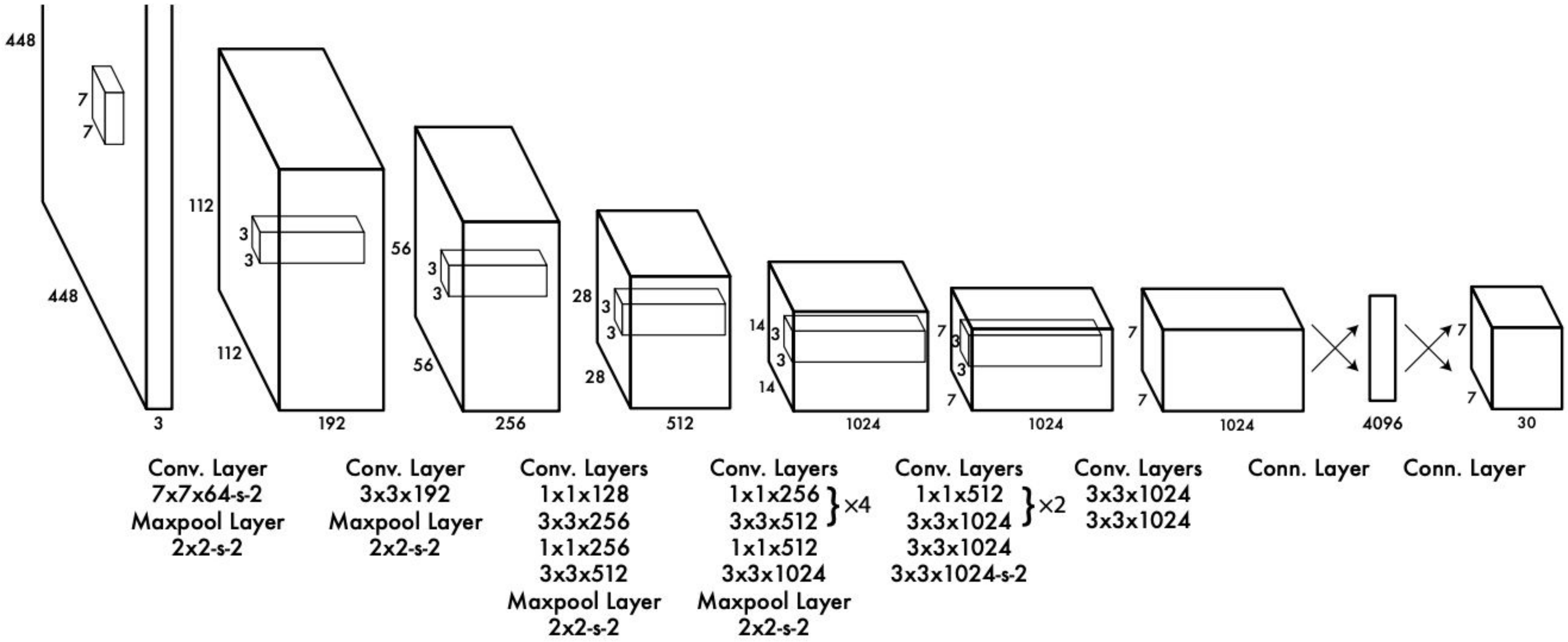
```yaml
# train_val_test.yaml
path: /datasets/coco8
train: images/train
val: images/val
test: images/test
# Classes
names:
  0: person
  1: bicycle
  2: car
```

Predict by Ultralytics

```
from ultralytics import YOLO

# Load a model

model = YOLO("yolo11n.pt")  # Pretrained YOLO11n model

# Run batched inference on a list of images

results = model(["image1.jpg", "image2.jpg"])  # Return a list of Results objects

# Process results list

for result in results:

    boxes = result.boxes  # Boxes object for bounding box outputs

    probs = result.probs  # Probs object for classification outputs

    result.save(filename="result.jpg")  # Save to disk
```

Architecture of YOLOv1

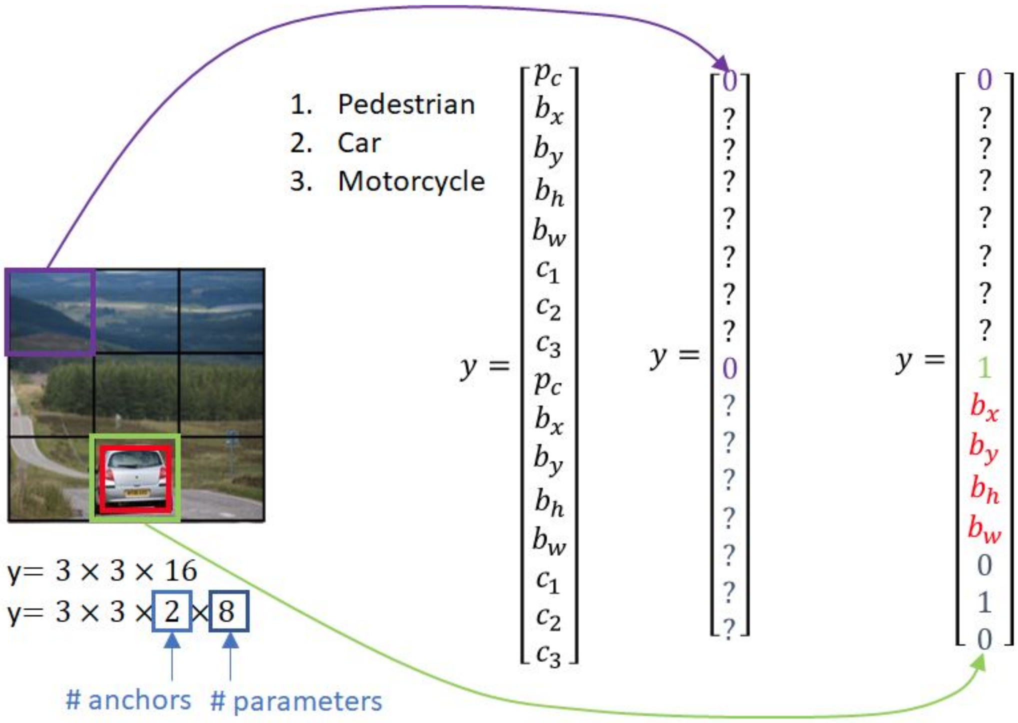
- 24 Convolution Layers + 4 MaxPooling Layers + 2 Fully Connected Layers



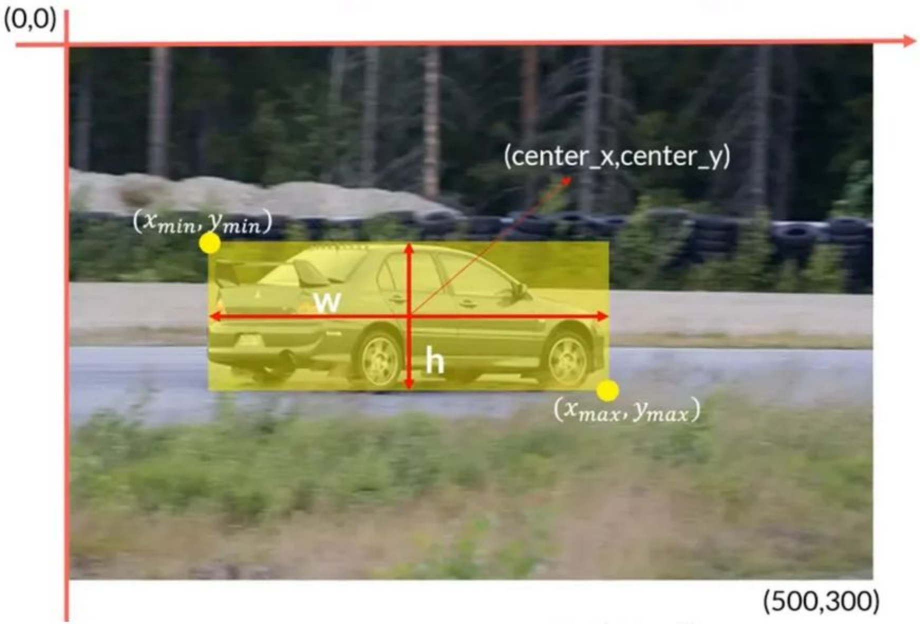
Output Vector of YOLO Model

- 3x3 grids
- 2 anchor boxes
- 8 parameters:

  $(p_c, b_x, b_y, b_w, b_h, c_0, c_1, c_2)$



Label Preparation for YOLO



- $center\_x = \dfrac{\frac{x_{min}+x_{max}}{2}}{width\ of\ the\ image}$
- $center\_y = \dfrac{\frac{y_{min}+y_{max}}{2}}{height\ of\ the\ image}$
- $w = \dfrac{x_{max} - x_{min}}{width\ of\ the\ image}$
- $h = \dfrac{y_{max} - y_{min}}{height\ of\ the\ image}$

Loss Function of YOLOv1

- Let sxs be total number of grid cells, then:

  Total Loss, $\mathrm{L} = \sum_{\mathrm{i}=1}^{\mathrm{s}^2} \mathrm{L_i}$

- Putting different weightage on different grid cells depending on whether they have objects and or not, total loss can be estimated:

$$\boldsymbol{L} = \sum_{i=1}^{S^2} \boxed{1_i^{obj}} \times L_{i,obj} + \lambda_{no\_obj} \sum_{i=1}^{S^2} \boxed{1_i^{no\_obj}} \times L_{i,no\_obj}$$

$1_i^{obj} = 1$ if i^{th} grid is **object anchor**

$1_i^{no\_obj} = 1$ if i^{th} grid is **no-object anchor**

Loss of Grid Cells Having Objects

- Loss of object grid = localization_loss + objectness_loss + classification_loss
- Putting more weightage on the localization_loss, loss for a single grid having an object: $L_{i,obj} = \lambda_{coord} \times L^{box}_{i,obj} + L^{conf}_{i,obj} + L^{cls}_{i,obj}$
- objectness_loss, $L^{conf}_{i,obj} = \left(c^*_i - \hat{c}_i\right)^2$
- classification_loss, $L^{cls}_{i,obj} = \sum_{c=1}^{n} \left(p^*_{i,c} - \widehat{p}_{i,c}\right)^2$
- When number of classes = 20, then classification_loss:

$$L^{cls}_{i,obj} = \sum_{c=1}^{20} \left(p^*_{i,c} - \widehat{p}_{i,c}\right)^2$$

Localization Loss or Bounding Box Regression Loss

Let

- Δx: normalized center_x
- Δy: normalized center_y, Δw: normalized width of the bounding box
- Δh: normalized height of the bounding box.

$$L_{i,obj}^{box} = \left(\Delta x_i^* - \Delta\hat{x}_i\right)^2 + \left(\Delta y_i^* - \Delta\hat{y}_i\right)^2 + \left(\sqrt{\Delta w_i^*} - \sqrt{\Delta\widehat{w}_i}\right)^2 + \left(\sqrt{\Delta h_i^*} - \sqrt{\Delta h_i}\right)^2$$

- $(\Delta\hat{x}_i, \Delta\hat{y}_i, \Delta\widehat{w}_i, \hat{h}_i)$: ground-truth box
- $(\Delta x_i^*, \Delta y_i^*, \Delta w_i^*, \Delta h_i^*)$: **responsible** predicted box that has the largest IoU with ground-truth box

Complete Loss Function of YOLOv1

Let:

- the number of classes, $n_c = 20$
- the number of grid cells = sxs

$$
\boldsymbol{L} = \lambda_{coord} \times \sum_{i=1}^{S^2} 1_i^{obj} \times \left(\begin{array}{c} (\Delta x_i^* - \Delta \hat{x}_i)^2 + (\Delta y_i^* - \Delta \hat{y}_i)^2 + \\ \left(\sqrt{\Delta w_i^*} - \sqrt{\Delta \widehat{w}_i}\right)^2 + \left(\sqrt{\Delta h_i^*} - \sqrt{\Delta h_i}\right)^2 \end{array} \right)
$$

$$
+ \sum_{i=1}^{S^2} 1_i^{obj} \times (c_i^* - \hat{c}_i)^2 + \sum_{i=1}^{S^2} 1_i^{obj} \times \sum_{c=1}^{20} \left(p_{i,c} - \hat{p}_{i,c}\right)^2
$$

$$
+ \lambda_{no\_obj} \sum_{i=1}^{S^2} 1_i^{no\_obj} \times \sum_{j=1}^{B} \left(c_{i,j} - \hat{c}_{i,j}\right)^2
$$