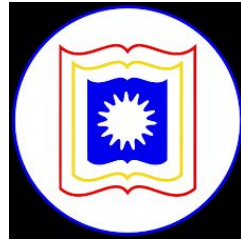


CSE4261: Neural Network and Deep Learning

Lecture: 01.07.2025



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

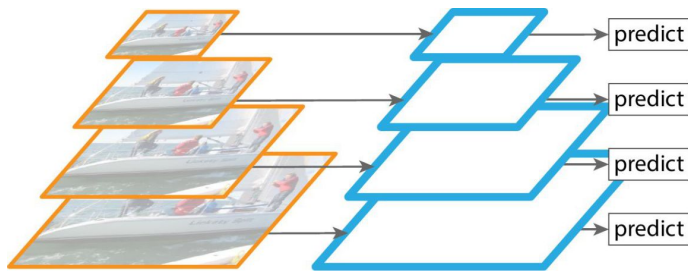
Different Versions of YOLO

1. YOLOv1 (2015)
 - Direct prediction of bounding boxes as a regression value
2. YOLOv2/YOLO9000 (2016)
 - Batch normalization, anchor boxes multi-scale predictions.
3. YOLOv3 (2018)
 - Darknet53 as feature extractor and multi-scale predictions by pyramidal feature hierarchy
4. YOLOv4 (2020)
 - Splitted model's architecture into three parts: Backbone, Neck, Head
5. YOLOv5 (2020)
6. YOLOv6 (2022)
7. YOLOv7 (2022)
8. YOLOv8 (2023)
 - Dynamic anchor-free detection
9. YOLOv9 (2024)
10. YOLOv10 (2024)
11. YOLOv11 (2024)
12. YOLOv12 (2025)

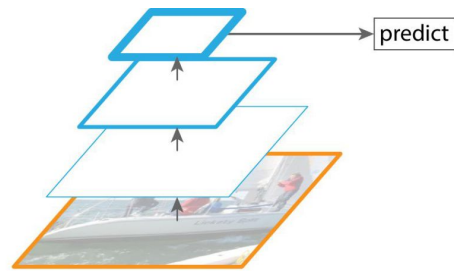
Attention-based models

Feature Pyramid Network (FPN) [2017]

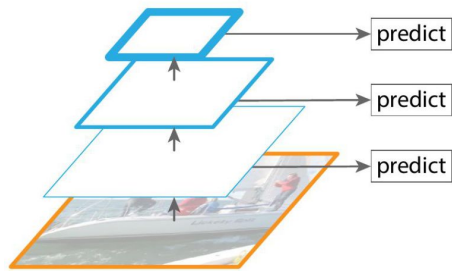
- It creates a feature pyramid from a single-scale input image.
- It enables object detection across various sizes.
- It combines high-resolution, low-level features with low-resolution, high-level semantic features.



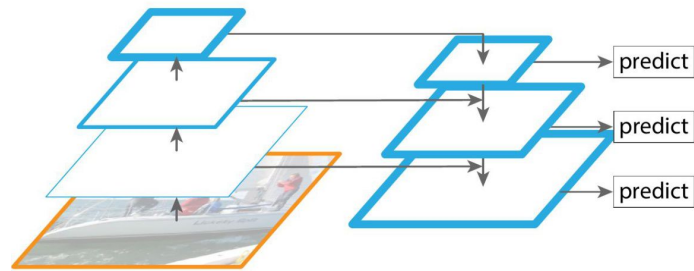
(a) Featurized image pyramid



(b) Single feature map

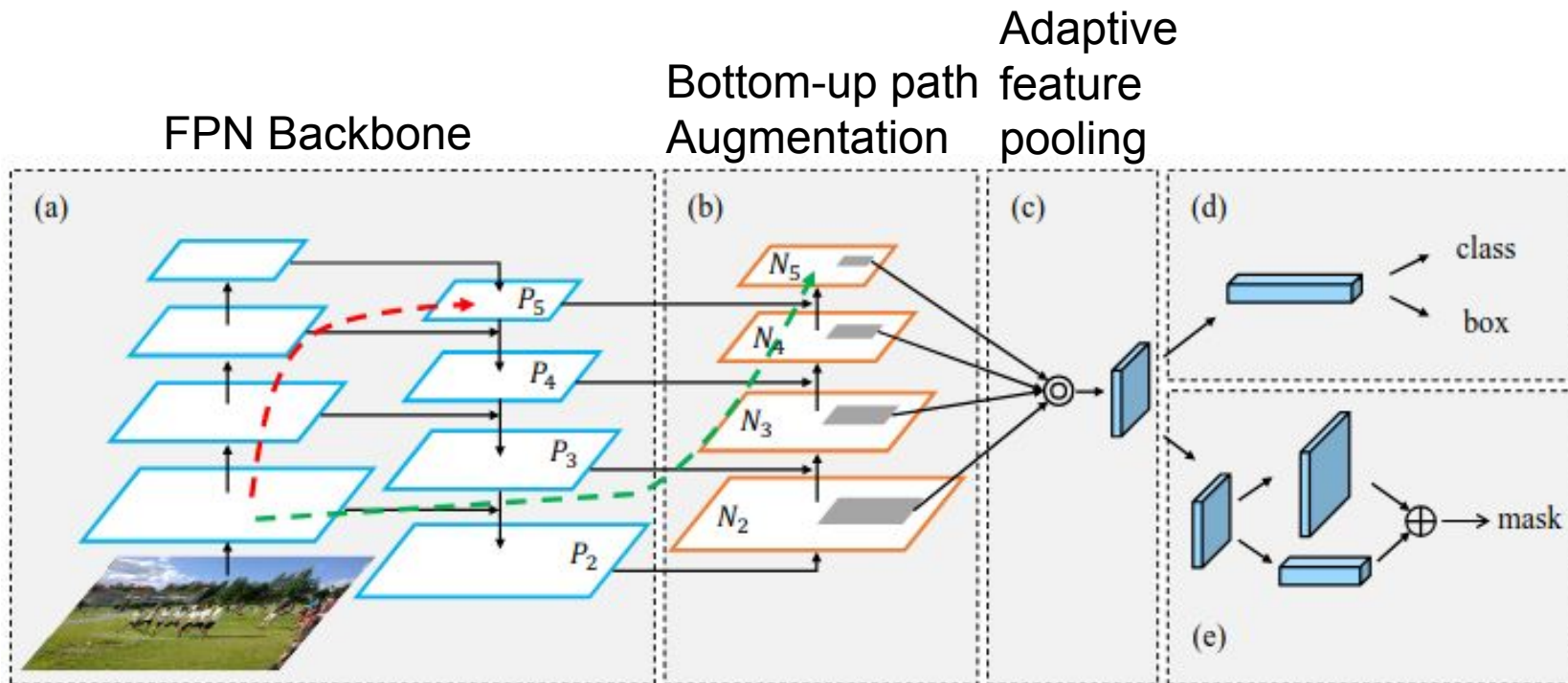


(c) Pyramidal feature hierarchy



(d) Feature Pyramid Network

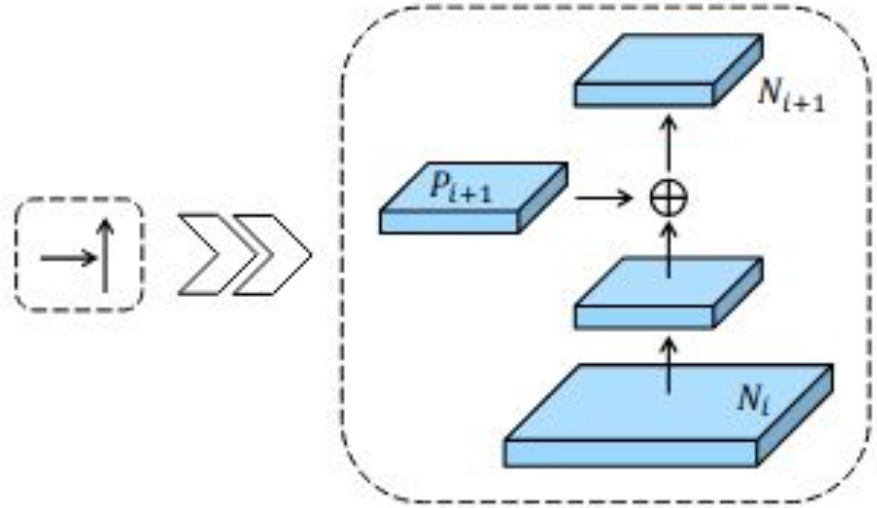
Path Aggregation Network (PAN/ PANet) [2018]



Building Block of Bottom-Up Path Augmentation

Each building block:

- takes a higher resolution feature map N_i and downsamples by Convolution with stride=2 to reduce spatial size
- adds with a coarser map P_{i+1} through lateral connection
- generates the new feature map N_{i+1}



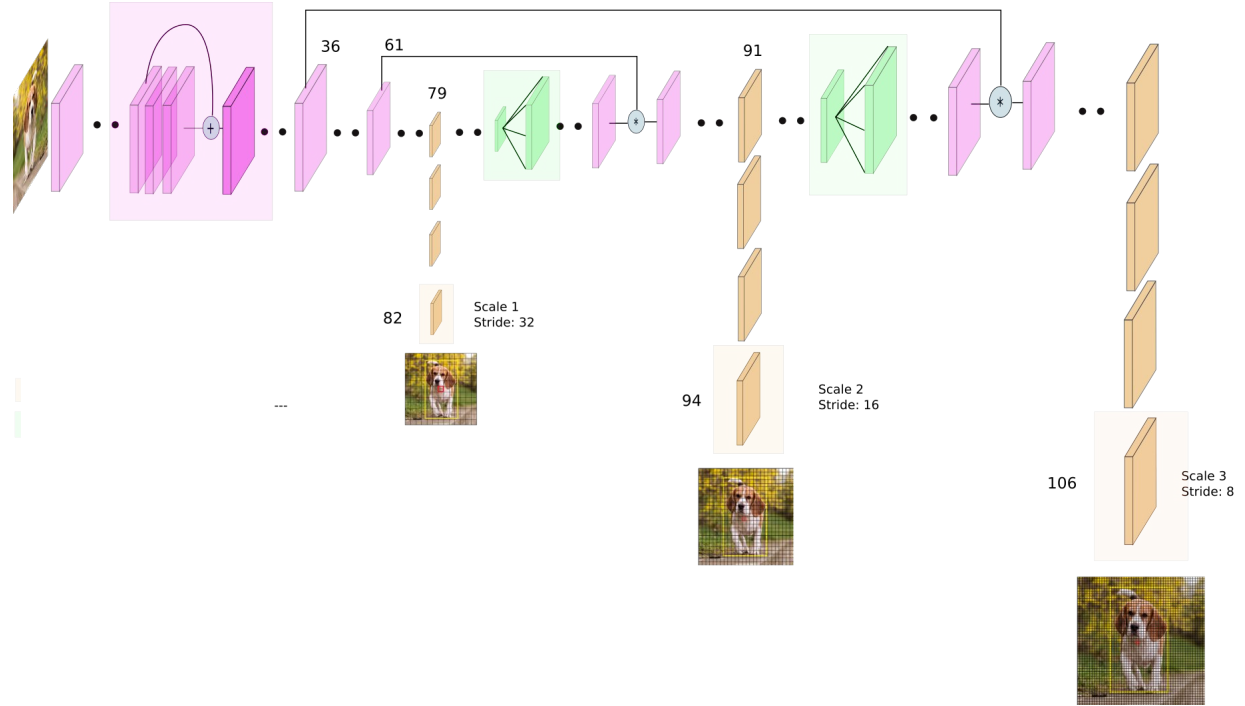
Architecture of YOLOv3 [2018]

- **Darknet53**

- It is a CNN with 53 layers.
- It utilizes 3x3 and 1x1 convolutional filters.
- It uses skip connection like ResNet

- **Note:**

- + : Addition
- * : Concatenation
- Both demand same size feature maps



Architecture of YOLOv4 [2020]

- Introduced the concept of the backbone, neck, and head architecture in YOLO

