BRAC UNIVERSITY

Inspiring Excellence

Object Detection YOLOv8

Submitted By: Group 4

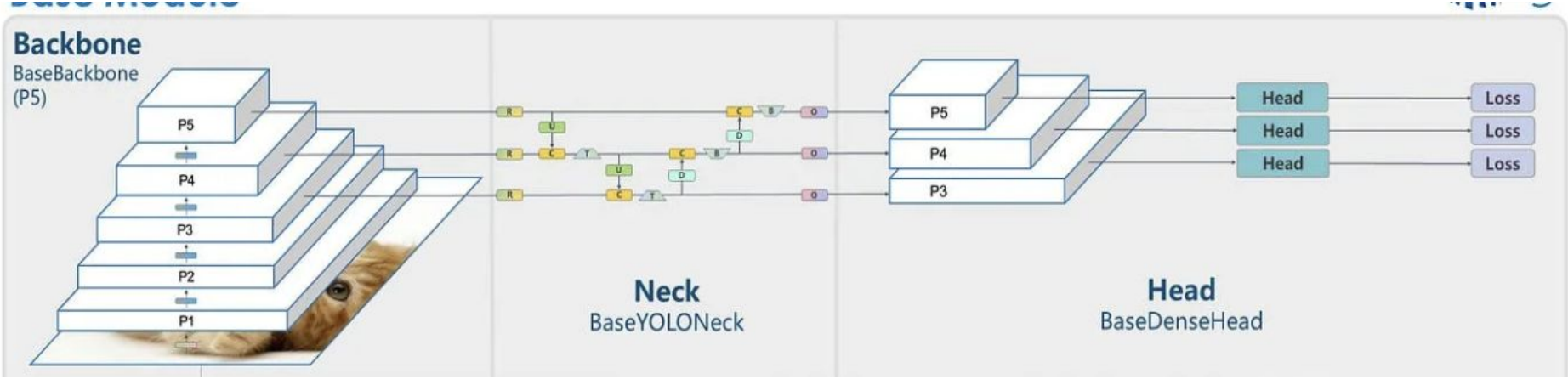
Samia Tasnim Orpa

Fahmida Akter

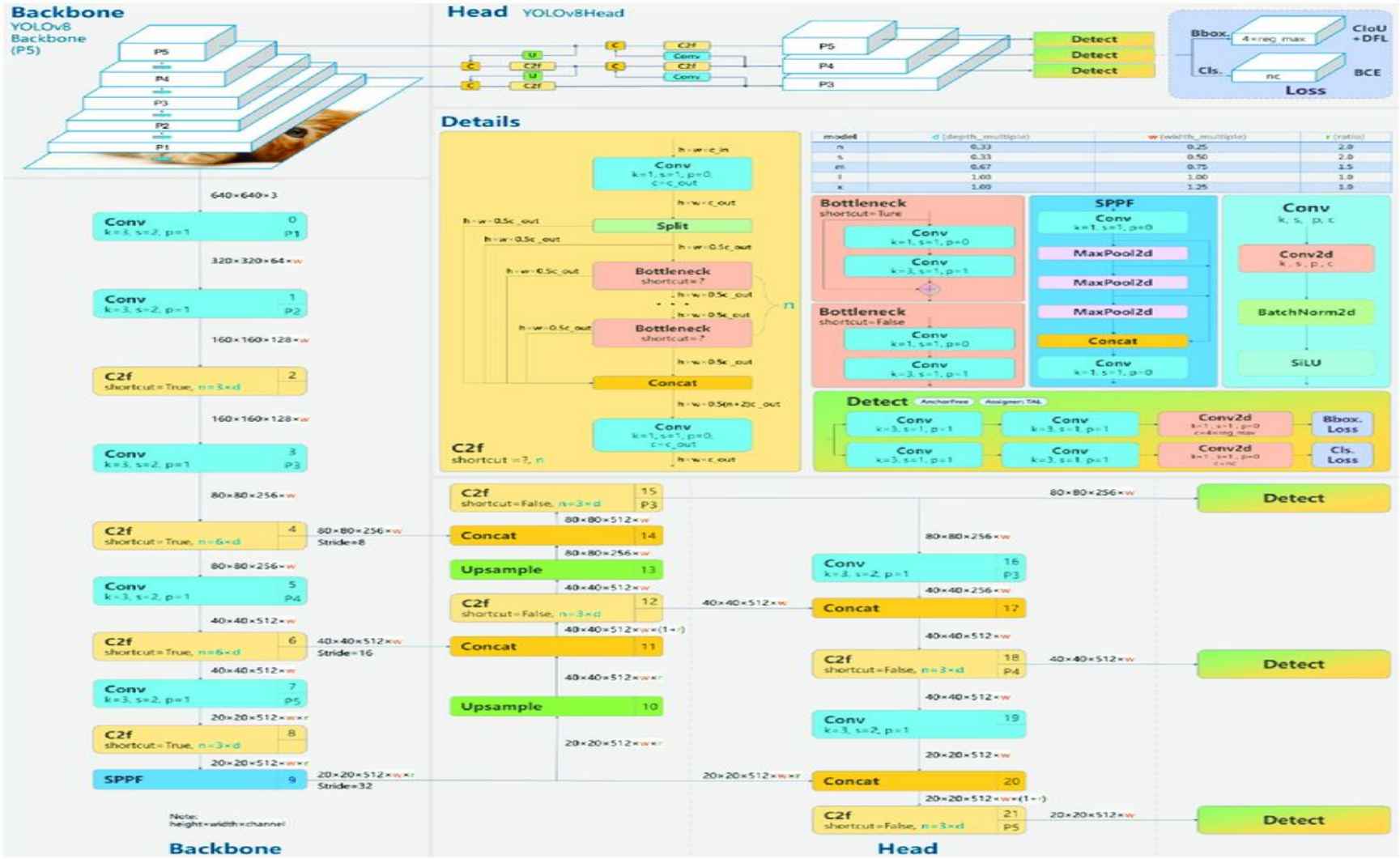
Nihat Rownok

YOLO v8 Model Architecture

YOLOv8 utilizes a Convolutional Neural Network that is divided into of 2 parts : The Backbone and The Head



- Backbone is the deep learning architecture that act as feature extractor
- In neck, the extracted features are pulled together and then they are put into the head.
- The head predicts the class and bounding box regions, which is the final output produced and are used as detection based on loss metrics.


Backbone
YOLOv8
Backbone
(P5)
P5
P4
P3
P2
P1
Head
YOLOv8Head
C
U
C2f
Conv
P5
P4
P3
Detect
Detect
Detect
Bbox.
4×reg_max
CIoU
+DFL
Cls.
nc
BCE
Loss
Details
h×w×c_in
Conv
k=1, s=1, p=0,
c=c_out
h×w×c_out
Split
h×w×0.5c_out
Bottleneck
shortcut=T
Concat
h×w×0.5(n+2)c_out
C2f
shortcut = T, n
n
model
d (depth_multiple)
w (width_multiple)
r (ratio)
n 0.33 0.25 2.0
s 0.33 0.50 2.0
m 0.67 0.75 1.5
l 1.00 1.00 1.0
x 1.00 1.25 1.0
Bottleneck
shortcut=Ture
Conv
k=1, s=1, p=0
Conv
k=3, s=1, p=1
Bottleneck
shortcut=False
SPPF
Conv
k=1, s=1, p=0
MaxPool2d
MaxPool2d
MaxPool2d
Concat
Conv
k=1, s=1, p=0
Conv
k, s, p, c
Conv2d
k, s, p, c
BatchNorm2d
SiLU
Detect
AnchorFree
Assigner: TAL
Conv
k=3, s=1, p=1
Conv2d
k=1, s=1, p=0
c=4×reg_max
Bbox.
Loss
Cls.
Loss
640×640×3
Conv
k=3, s=2, p=1
0
P1
320×320×64×w
Conv
k=3, s=2, p=1
1
P2
160×160×128×w
C2f
shortcut=True, n=3×d
2
160×160×128×w
Conv
k=3, s=2, p=1
3
P3
80×80×256×w
C2f
shortcut=True, n=6×d
4
80×80×256×w
Stride=8
80×80×256×w
Conv
k=3, s=2, p=1
5
P4
40×40×512×w
C2f
shortcut=True, n=6×d
6
40×40×512×w
Stride=16
40×40×512×w
Conv
k=3, s=2, p=1
7
P5
20×20×512×w×r
C2f
shortcut=True, n=3×d
8
20×20×512×w×r
SPPF
9
20×20×512×w×r
Stride=32
Note:
height×width×channel
Backbone
C2f
shortcut=False, n=3×d
15
P3
80×80×512×w
Concat
14
80×80×256×w
Upsample
13
40×40×512×w
C2f
shortcut=False, n=3×d
12
40×40×512×w×(1+r)
Concat
11
40×40×512×w×r
Upsample
10
20×20×512×w×r
80×80×256×w
Detect
80×80×256×w
Conv
k=3, s=2, p=1
16
P3
40×40×256×w
40×40×512×w
Concat
17
40×40×512×w
C2f
shortcut=False, n=3×d
18
P4
40×40×512×w
Detect
40×40×512×w
Conv
k=3, s=2, p=1
19
20×20×512×w
20×20×512×w×r
Concat
20
20×20×512×w×(1+r)
C2f
shortcut=False, n=3×d
21
P5
20×20×512×w
Detect
Head

Features:

- One of the key features of YOLOv8 is the use of a **self-attention** mechanism in the head of the network.
- This mechanism allows the model to focus on different parts of the image and adjust different features based on their needs.
- Another important feature of YOLOv8 is its ability to perform **multi-scaled object detection**. It uses feature pyramid network consists of multiple layers that detect objects at different scales, allowing the model to detect large and small objects.

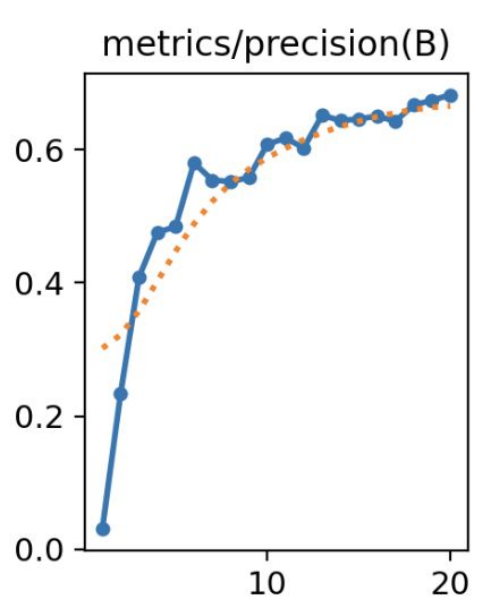
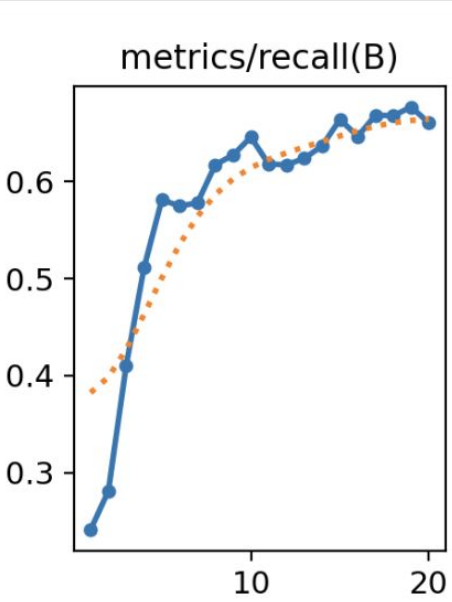
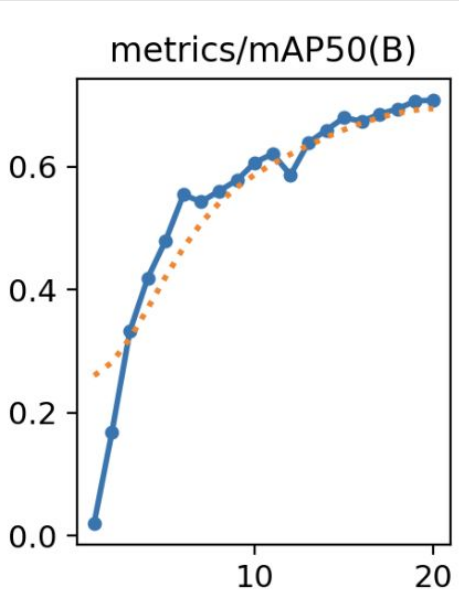
Metrics:

IoU **Precision** **Recall** **mAP50**

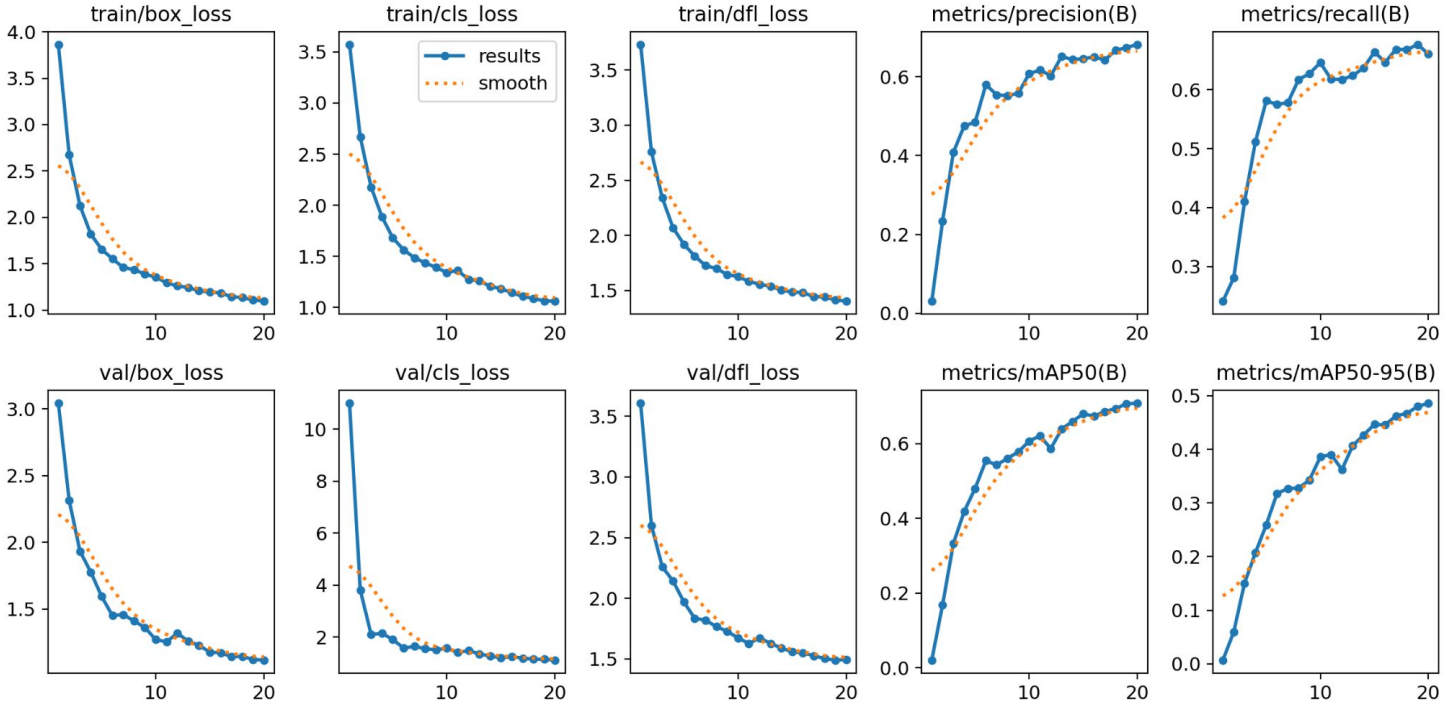
Metrices

	Train (after 20 epochs)	Validation	Test
Precision	0.683	0.682	0.713
Recall	0.662	0.66	0.699
IoU threshold (set by user)	0.7	0.6	0.7
mAP50	0.71	0.708	0.735

Metrices (Training for 20 epochs)

Precision	Recall	mAP50
		

Training Results Graphs



Metrices (Validation)

For the iou threshold = 0.6, the validation metrics are:

Class	Images	Instances	Box(P	R	mAP50	mAP50-95):
all	713	1602	0.682	0.66	0.708	0.486

Testing Phase

IoU threshold = 0.7
Confidence threshold = 0.25

Bad (Test) Detection Example 2

- Low confidence level
- Large part of building is detected instead of a signboard
- 2 signboards in image, 1 not detected at all, other detected with whole building

More Bad (Test) Detection Example

Contribution List

References

1. https://docs.ultralytics.com/guides/yolo-performance-metrics/#introduction
2. https://docs.ultralytics.com/modes/
3. https://blog.roboflow.com/whats-new-in-yolov8/
4. https://youtu.be/m9fH9OWn8YM?si=ZV2i0ARdJX0OMDHD