

COMPARISON OF OBJECT DETECTION ARCHITECTURAL MODELS

DEEPMLEARNING PROJECT PRESENTATION

07 Sep, 2024

LIST OF ARCHITECTURES

R-CNN: Regions with Convolutional Neural Networks

YOLO: You Only Look Once

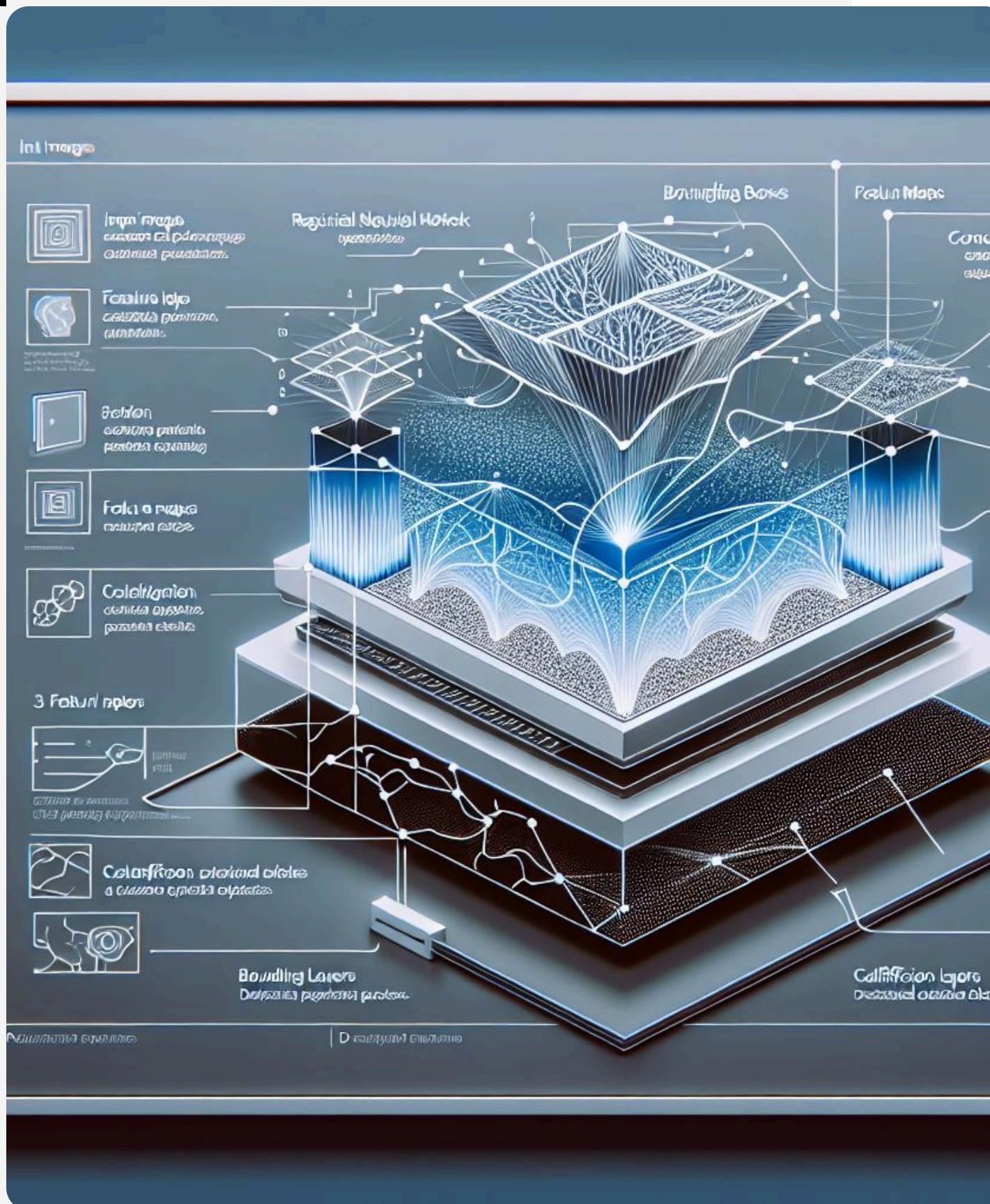
SSD: Single Shot MultiBox Detector



DATASET

- **Cabbages:** 233 from one variety of all cabbage.
 - **Train:** 173
 - **Validate:** 35
 - **Test:** 25
- **Size:** 640 x 640 px.
- **Label process:**  Roboflow (Export vary format).

METHODOLOGY



FIXED HYPERPARAMETERS

Epochs: 50

Batch size: 16

Other Configuration: Utilized transfer learning from a pretrained model to achieve optimal performance on designed architectures.

COMPARE THE ARCHITECTURES WITH

1. **mAP:** Mean Average Precision in %.
2. **Inference Time:** Prediction time in ms.
3. **Model size:** Efficiency per metrics.
4. **Optional:**
 - a. Frame Rate: $\text{FPS} = 1000/\text{Inference Time(ms)}$.
 - b. Training Time: Very little impact in this project.

R-CNN

Regions with Convolutional Neural Networks

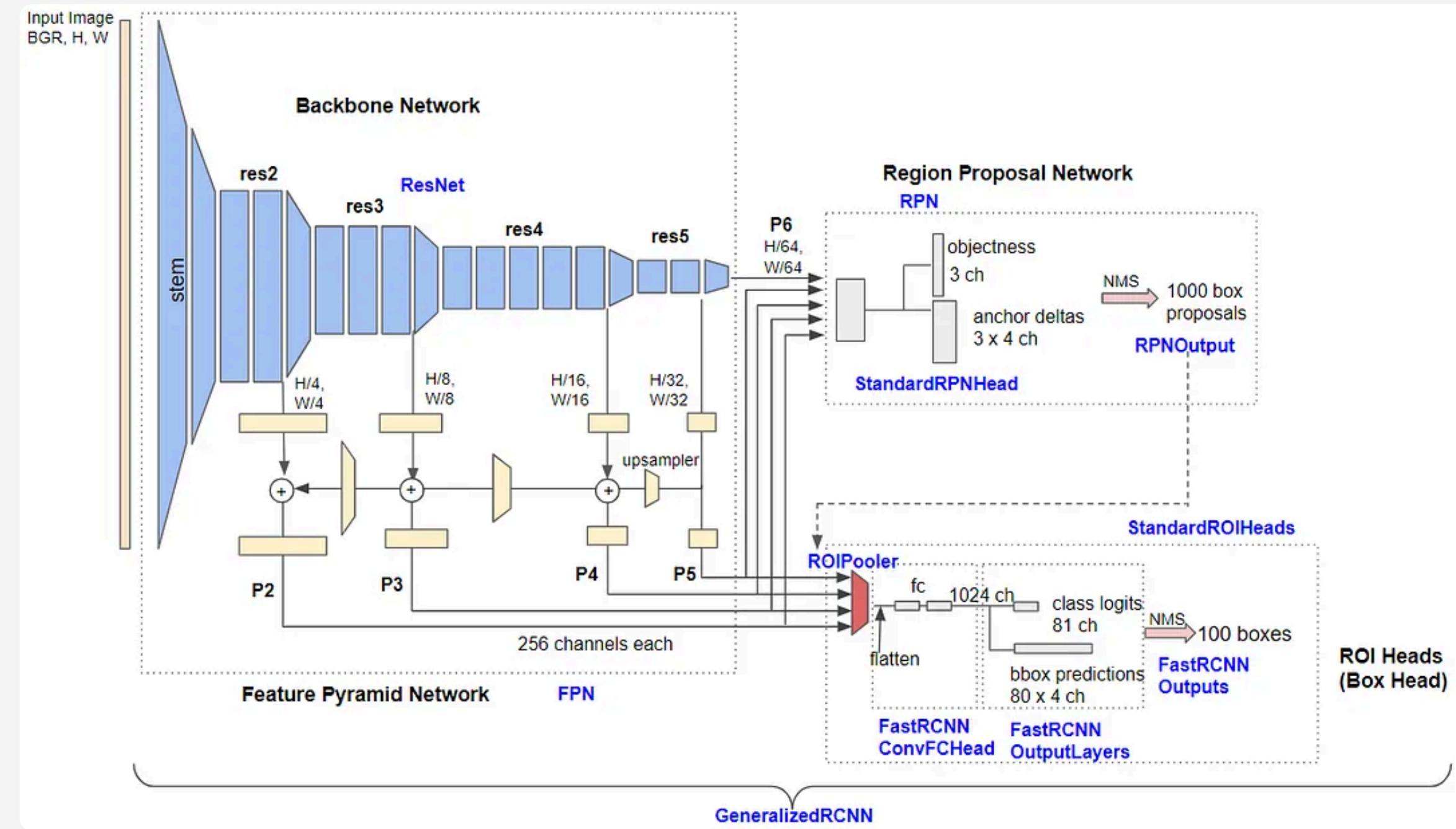
Two-stage object detectors



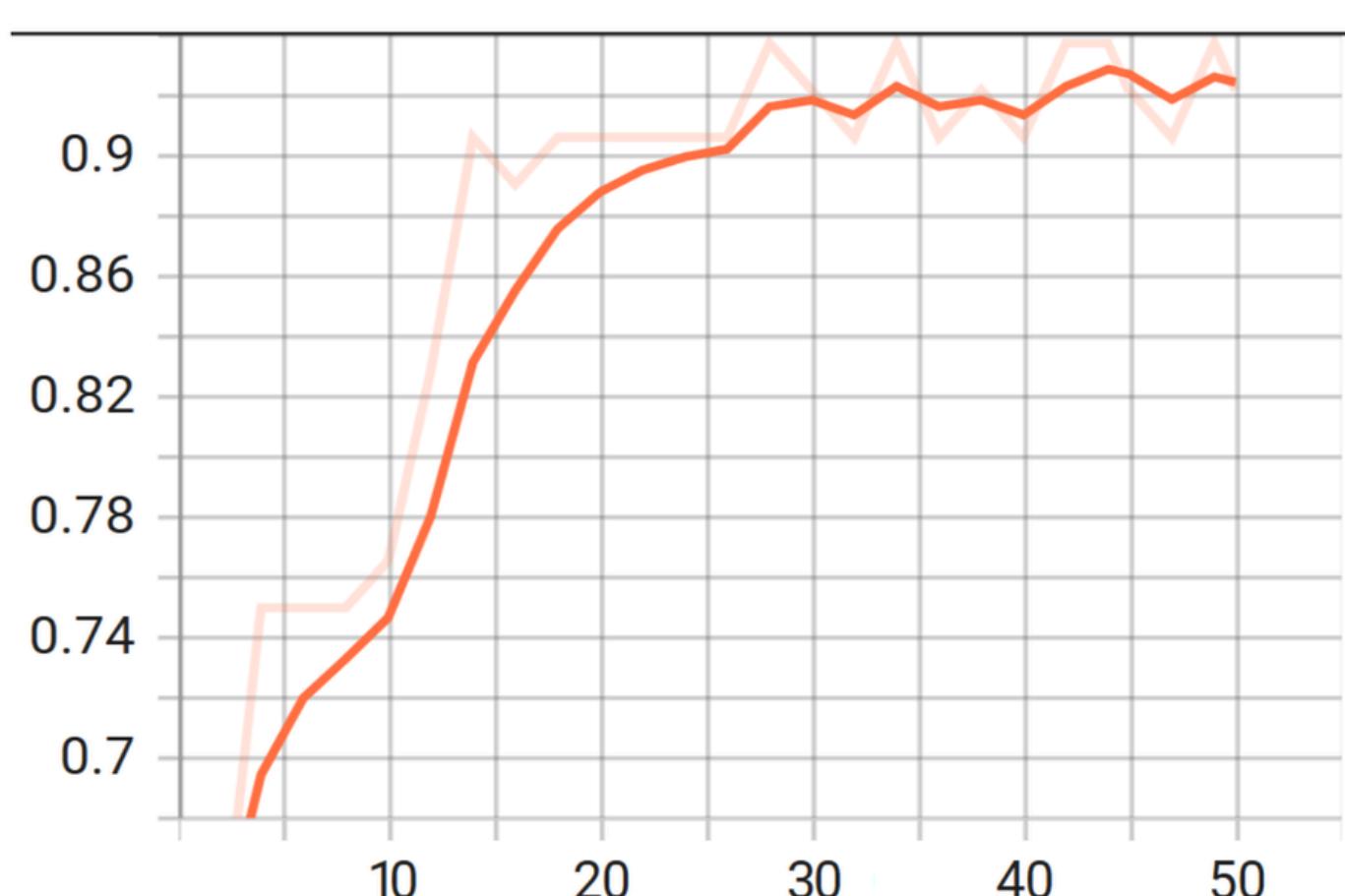
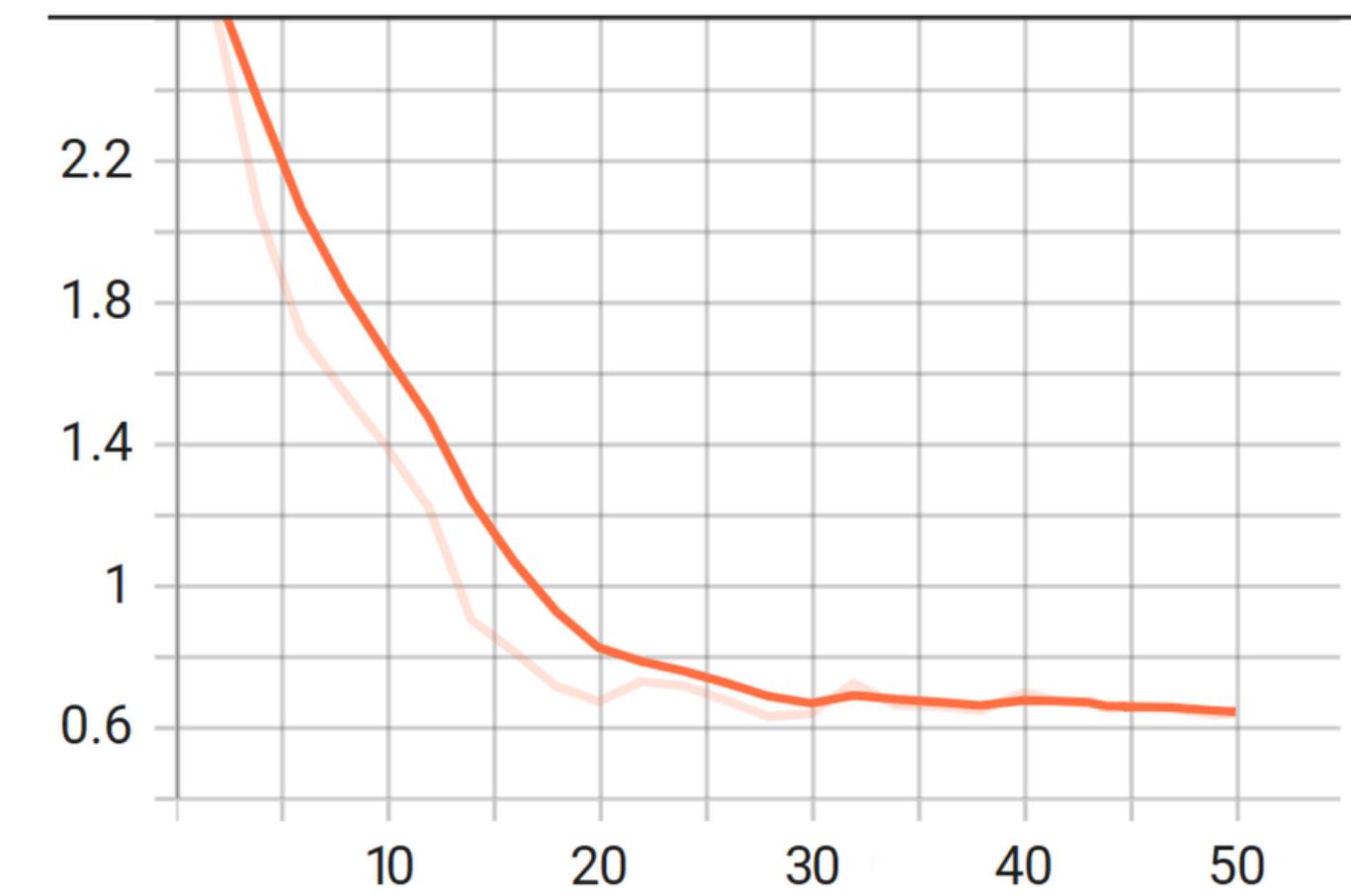
Detectron2

ResNet101 FNF 3x
with Pytorch

STRUCTURE OF THE BASE R-CNN FPN



Total Loss



TRAINING AND VALIDATION

- Training time: 18 mins or 1092 secs
- mPA : 96.52%

PREDICTION

Inference Time: ~72 ms (~13.89 FPS)



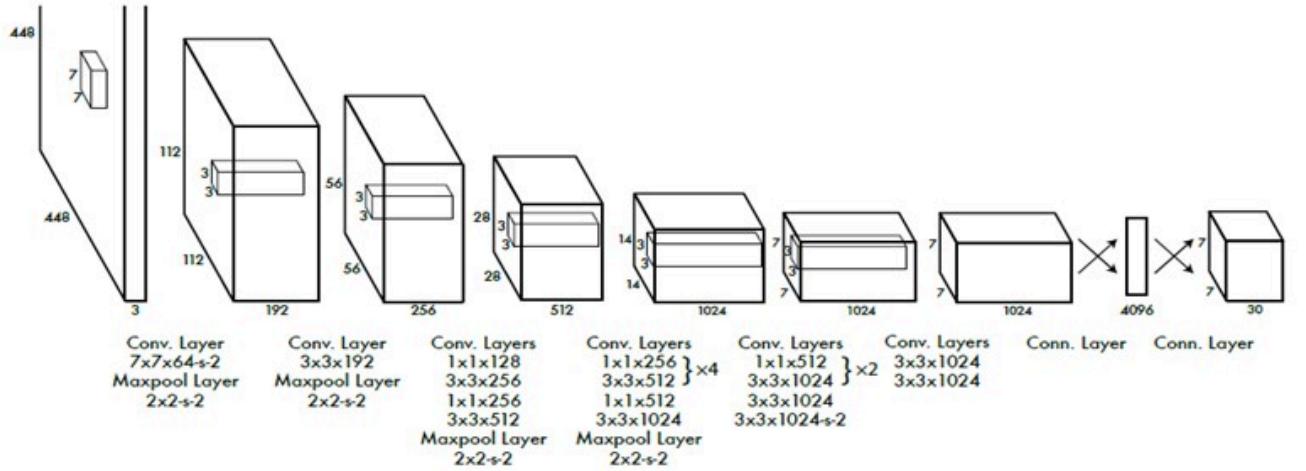


YOLO
You Only Look Once

Single-stage object detectors

YOLOV11

STRUCTURE OF YOLO

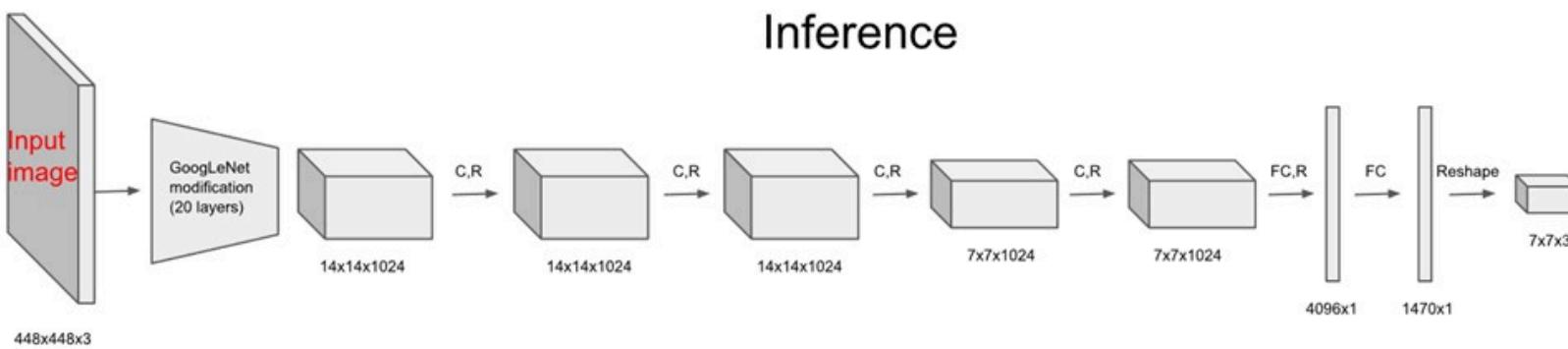


$$\text{Leaky ReLU} \\ \phi(x) = \begin{cases} x, & \text{if } x > 0 \\ 0.1x, & \text{otherwise} \end{cases}$$



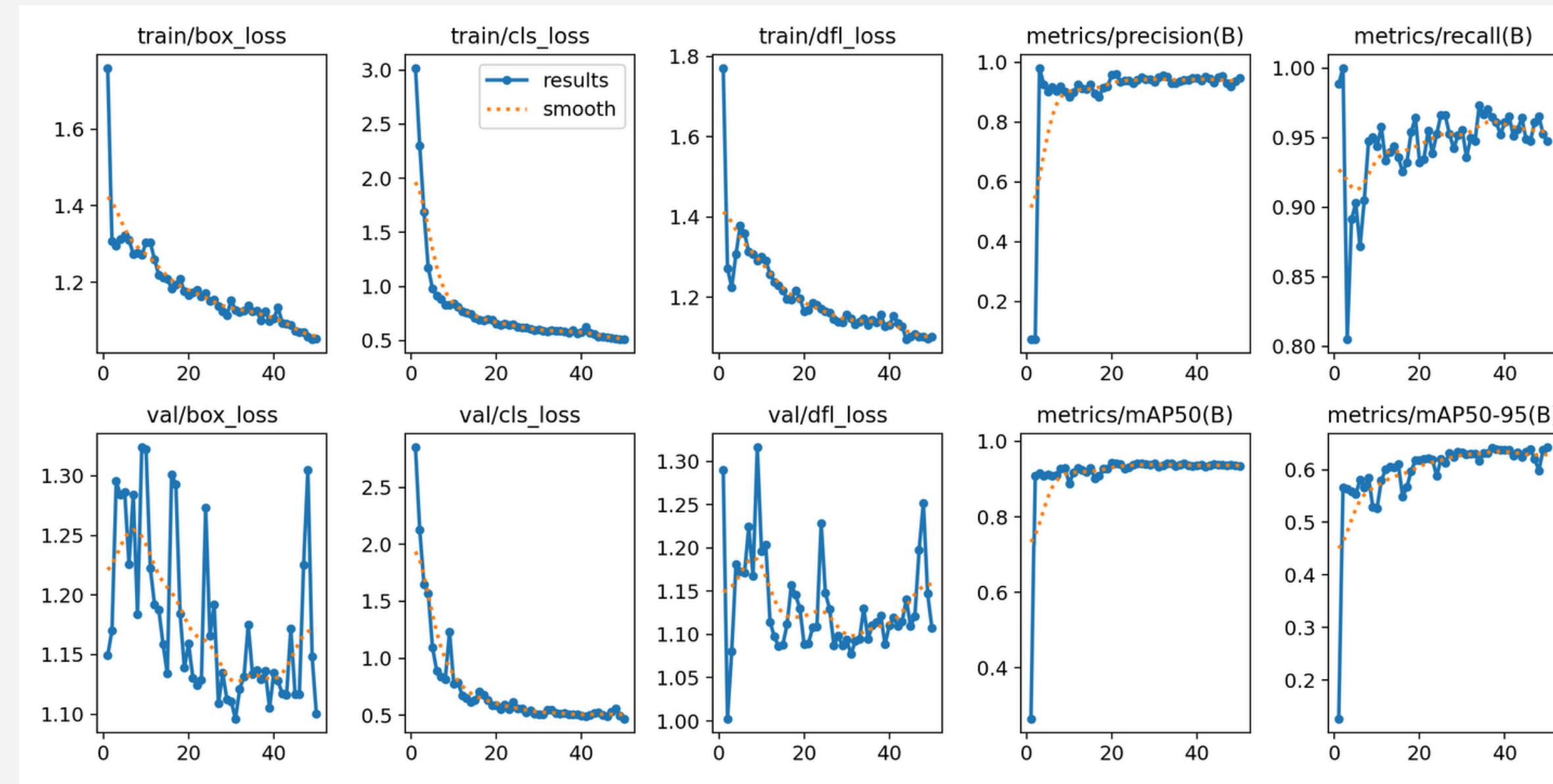
LINEAR ACTIVATION FUNCTION
(last layer)

YOLOv11n (nano model)



TRAINING AND VALIDATION

- Training time: 5.1 mins or 306 secs
- mPA : 91.17%



PREDICTION

Inference Time: ~4.7 ms (~212.76 FPS)



SSD

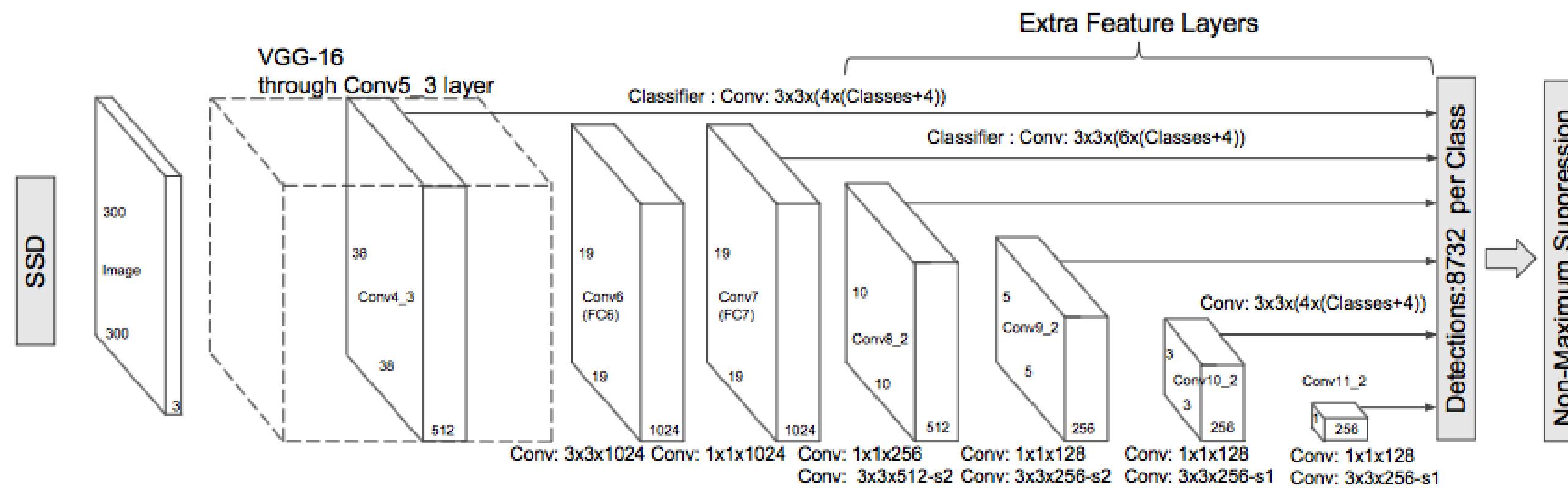
Single Shot MultiBox Detector

Single-stage object detectors

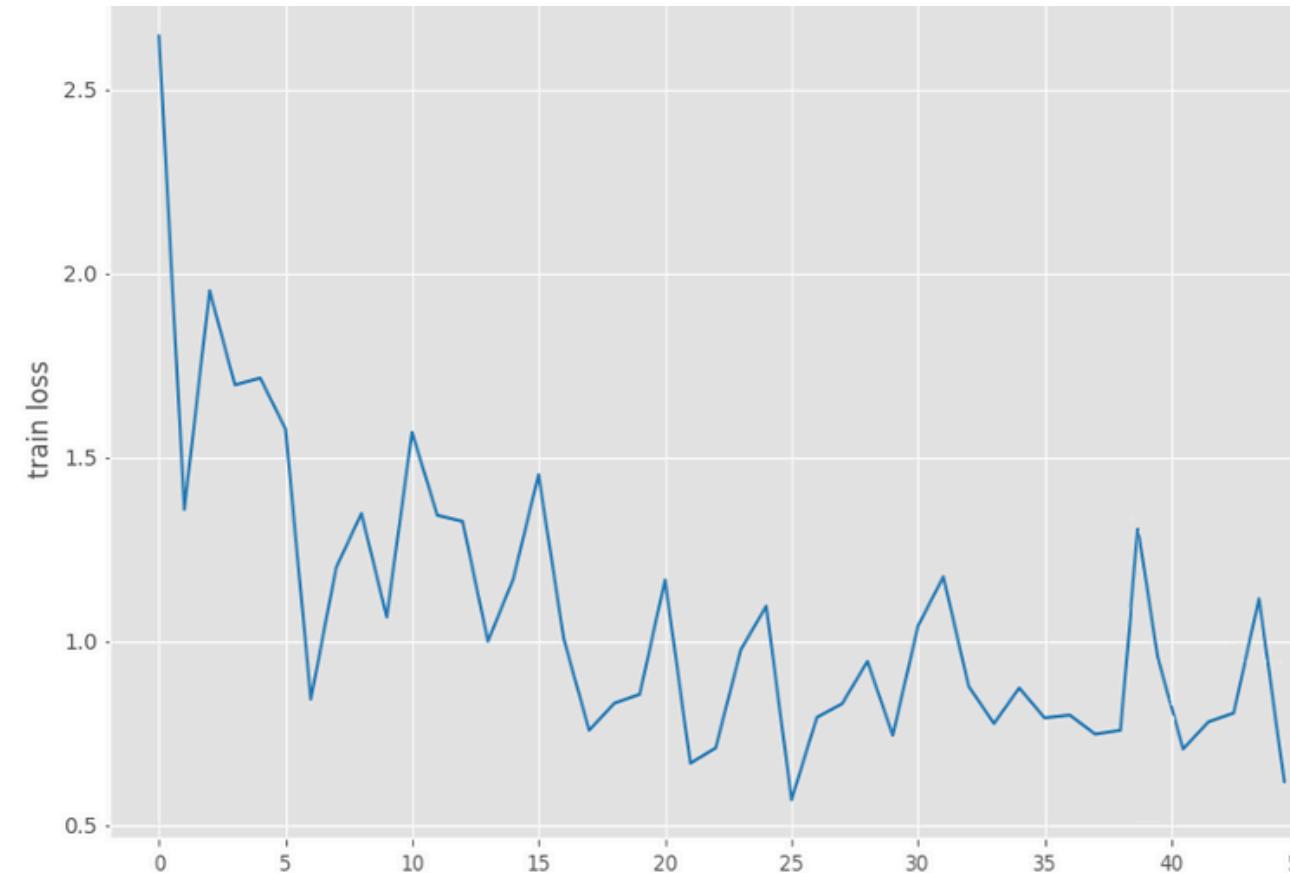
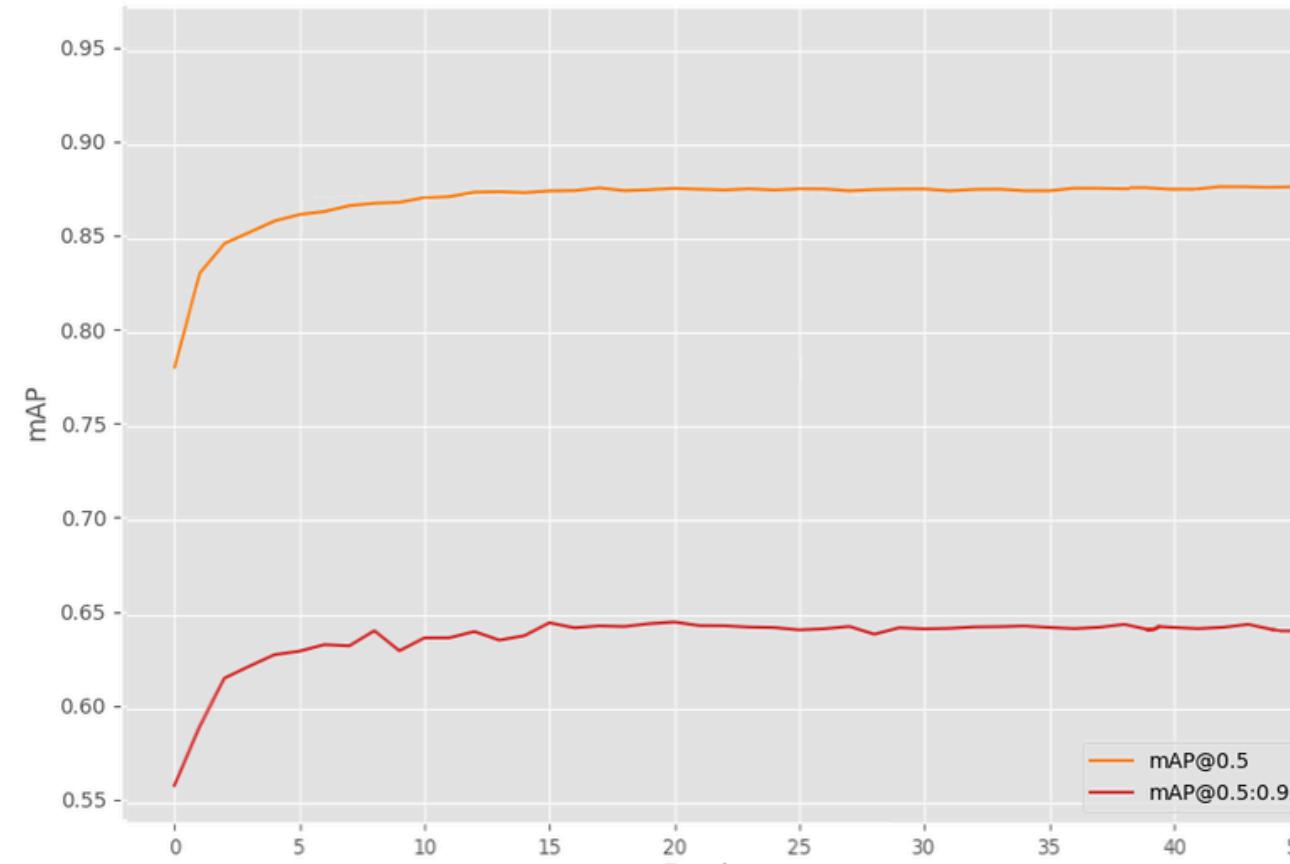
SSD

VGG-16 Backbone

STRUCTURE OF SSD



TRAINING AND VALIDATION



Training time: 13 mins or 306 secs
mPA : 87.41%

PREDICTION

Inference Time: ~55 ms (~18.18 FPS)



CONCLUSION

COMPARISON



1. Faster R-CNN: very complex due to two stage architecture.
2. SSD: like Faster R-CNN but in single stage, so less complexity.
3. YOLO: simple and small.

1. Faster R-CNN: ~97.52%
2. SSD: ~91.41%
3. YOLO: ~91.17%

1. YOLO: ~4.7 ms
2. SSD: ~55 ms
3. Faster R-CNN: ~72 ms

1. YOLO: 4.7 MB
2. SSD: 113 MB
3. Faster R-CNN: 256 MB

OPTIONAL:

- TRAINING TIME:
 - YOLO: ~5.1 MINS
 - SSD: ~16 MINS
 - Faster R-CNN: ~18 mins
- [ON 173 TRAINING DATA]

- FRAME RATE:
 - YOLO: ~212.76 FPS
 - SSD: ~18.18 FPS
 - Faster R-CNN: ~13.89 FPS
- [BASED ON INFERENCE TIME]

APPLICATIONS USES

REAL-TIME APPLICATIONS

YOLO

(Low inference time high frame rate required for real-time applications)

HIGH ACCURACY APPLICATIONS

R-CNN

(Highest mPA that necessary for high accuracy needed applications)

BALANCE NEEDED APPLICATIONS

SSD

(Balance between accuracy and speed optional for the needed balance application)