



REPORT DEEP LEARNING

Performance Comparison of YOLOv5 and YOLOv8 Architectures

Team :

NGUYEN Anh Tuan
LE Doan Phuoc
NGUYEN Ke An
PHAM Tien Duy

Professor :

Mr. Khaled AL-SAIH

University year : 2023 – 2024

1. Introduction

- **About the datasets** : a dataset about the image with 80 different classes. The dataset has 17648 images, in which the train set has 15466 images (88%), the validation set has 1290 images (7%) and the test set has 892 images (5%). We also take some preprocessing and data augmentations like this :

Preprocessing

Auto-Orient: Applied

Resize: stretch to 640x640

Augmentations

Outputs per training example: 2

Rotation: Between -45° and $+45^{\circ}$

Noise: Up to 0.02% of pixels

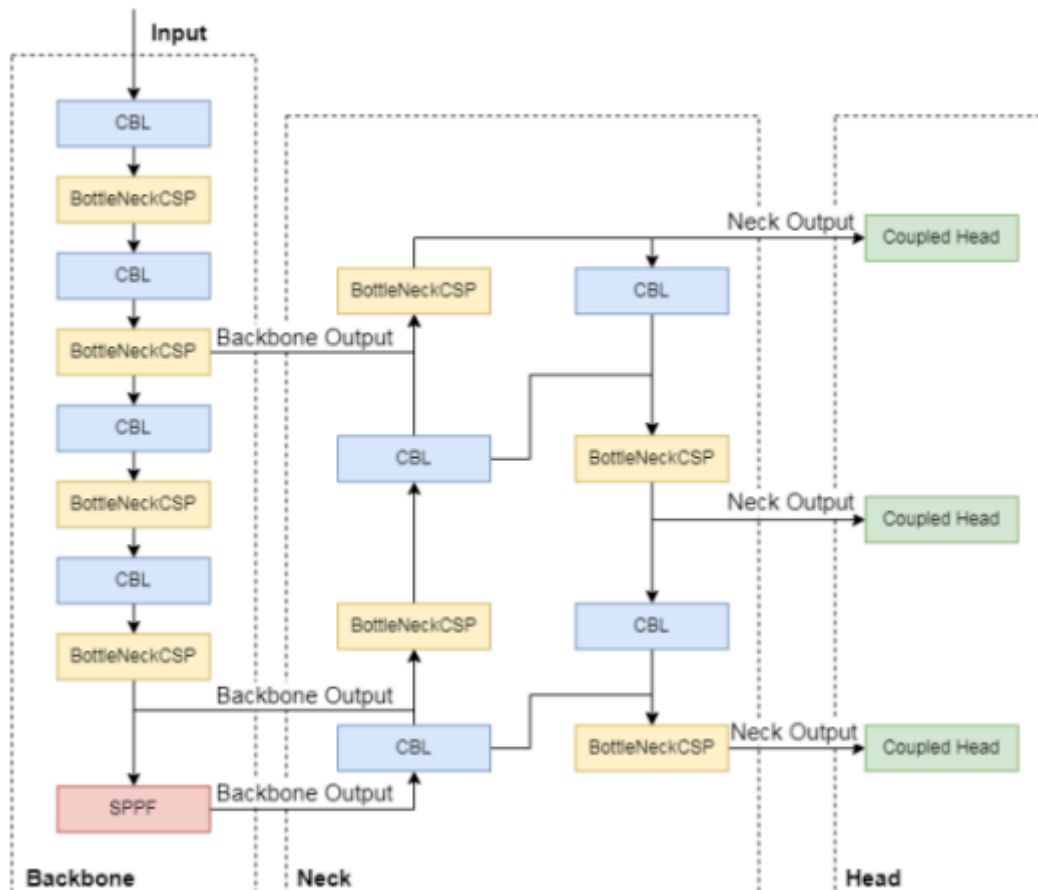
- **About architecture** : We're going to use both architecture YOLOv5 and YOLOv8 to deal with the datasets and compare the performance.
- **About evaluation** : The training results were then used to evaluate 15466 images over 10 epochs. The evaluation metric used was mAP which is equal to the average of the Average Precision metric across all classes in a model, the Precision and Recall and confusion metric.

2. Difference between YOLOv5 and YOLOv8

A. YOLOv5

The model YOLOv5's arrangement is divided into three components: backbone, neck, and head. The backbone of YOLOv5 is Darknet 53. Darknet 53 is a new network architecture that focuses on feature extraction characterized by small filter windows and residual connections. The neck of YOLOv5 acts as a connector between the backbone and the head. The neck of YOLOv5 functions to gather and refine the features extracted by the backbone, with a focus on enhancing spatial and semantic information at various scales. The head of YOLOv5 consists of three branches, each

predicting features at different scales. Each head produces bounding boxes, class probabilities, and confidence scores. Finally, the network uses Non-maximum Suppression (NMS) to filter overlapping bounding boxes. Here is the architecture of YOLOv5 :

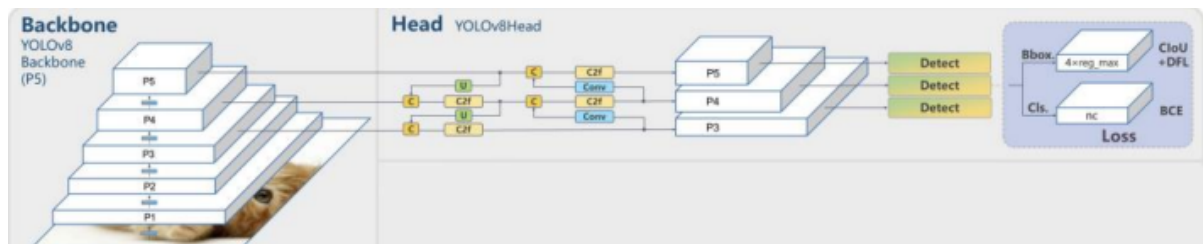


B. YOLOv8

YOLOv8 is the latest version of the object detection model architecture, succeeding YOLOv5. YOLOv8 introduces improvements in the form of a new neural network architecture. Two neural networks are implemented, namely the Feature Pyramid Network (FPN) and the Path Aggregation Network (PAN), along with a new labeling tool that simplifies the annotation process. This labeling tool contains several useful features, such as automatic labeling, shortcut labeling, and customizable hotkeys. The combination of these features makes it easier to annotate images for training the model.

FPN works by gradually reducing the spatial resolution of the input image while increasing the number of feature channels. This results in the creation of a feature map that is capable of detecting objects at different scales and resolutions. On the other hand, the PAN architecture can combine features from different network levels through skip connections. Consequently,

the network can capture features more effectively at various scales and resolutions, which is crucial for accurately detecting objects of different sizes and shapes. Here is the architecture of YOLOv8:

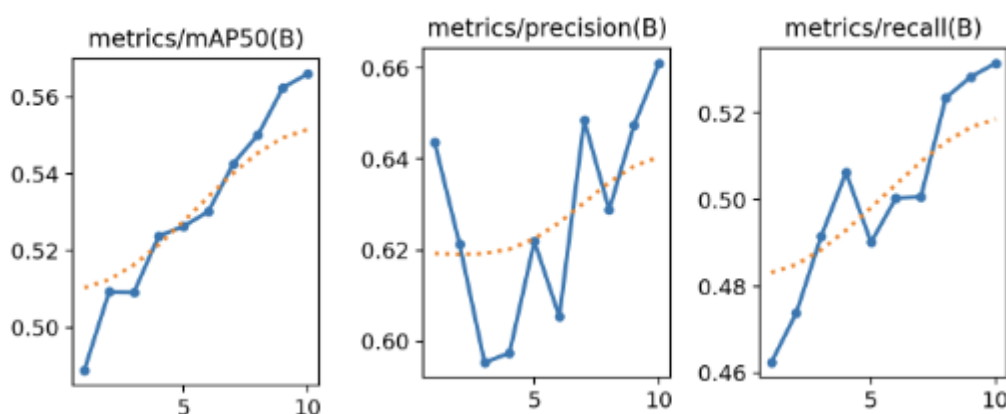


3. Comparison the performance of two architectures

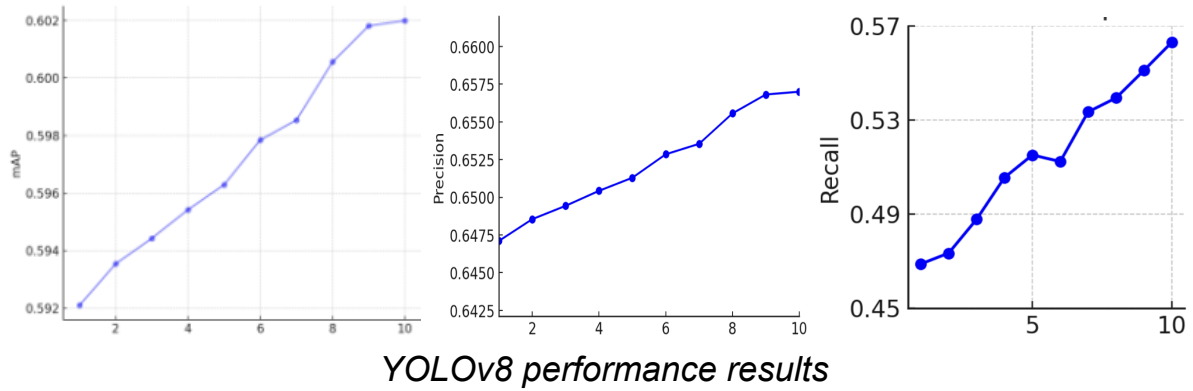
- Performance on the test set

Architecture	mAP	Precision	Recall
YOLOv5	55.6%	66.1%	53.2%
YOLOv8	60.2%	65.7%	56.3%

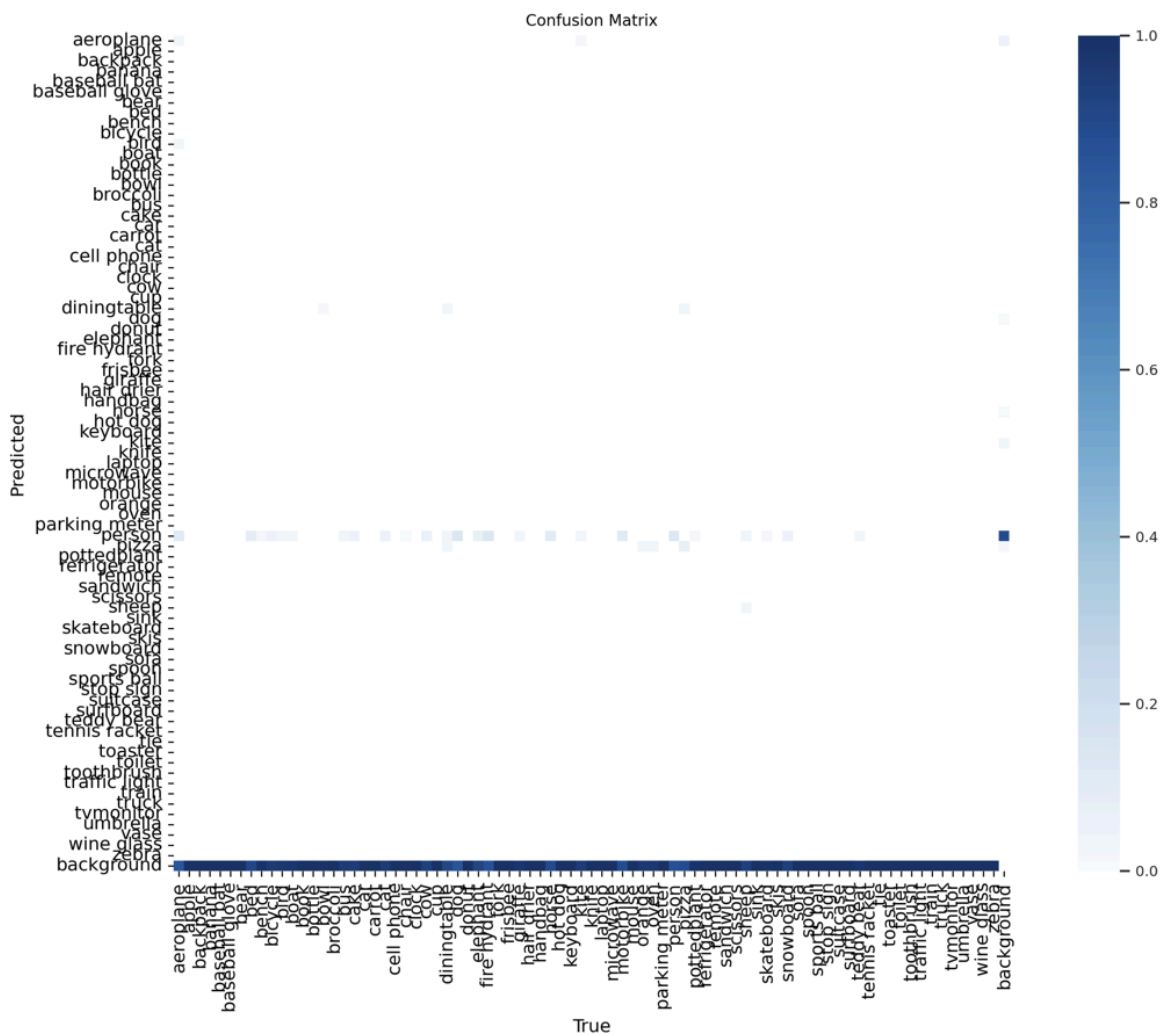
The performance values of the YOLOv5 and YOLOv8 models are summarized in the table above. The YOLOv8 model is 3.6% better than YOLOv5 for the mAP value. The difference in the precision value of the YOLOv5 model is slightly greater than that of YOLOv8 which is 0.4%. The recall value of the YOLOv8 model is greater than that of YOLOv5 by 3.1%.

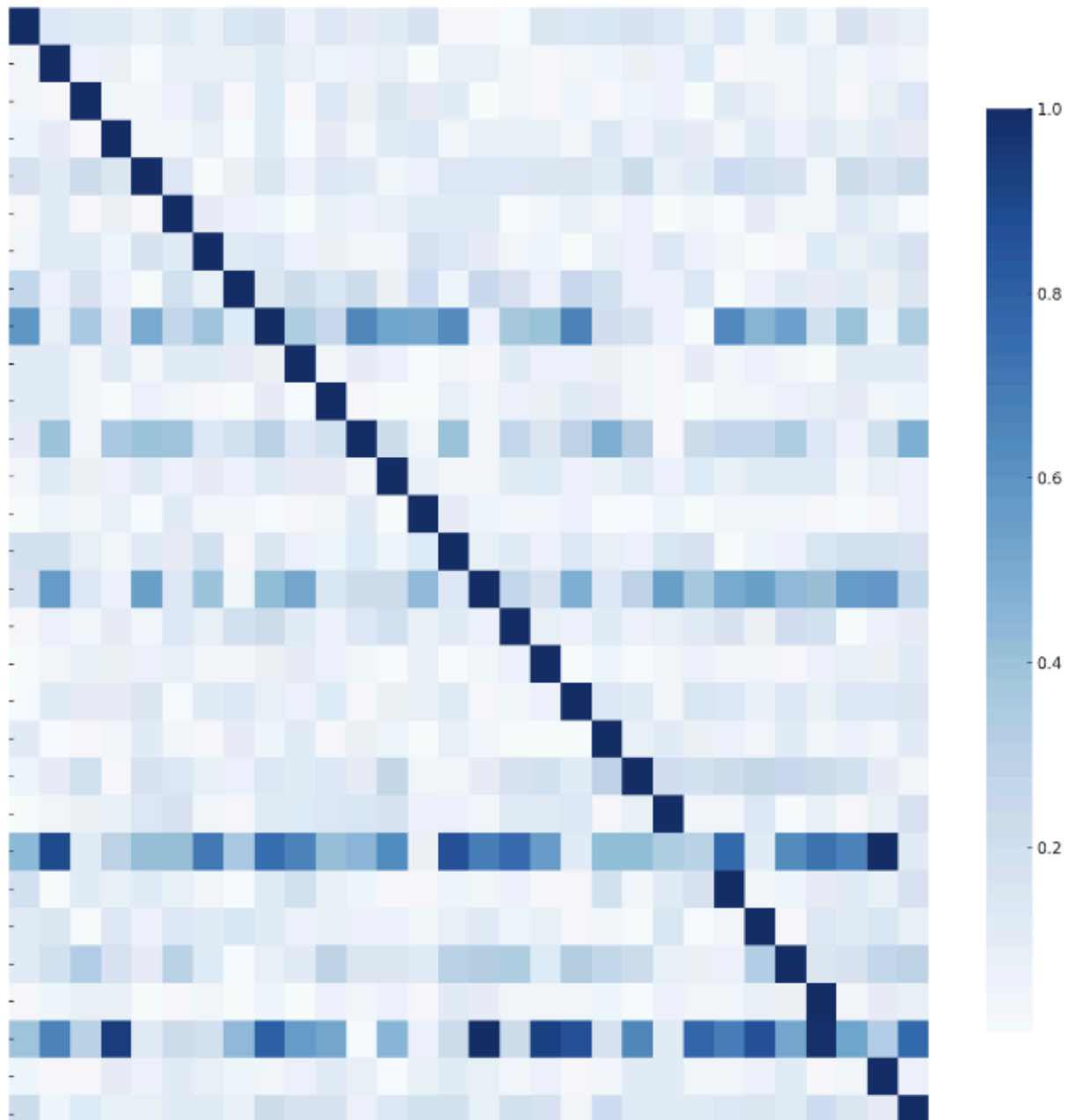


YOLOv5 performance results



- Confusion matrix:





YOLOv8's confusion matrix

4. Conclusions

Based on the research results discussed, it is known that the YOLOv5 and YOLOv8 models have successfully applied for the images in the dataset. There are differences in performance values but the performance value of the YOLOv8 model is greater than the YOLOv5 model for recall and mAP, the difference in the value of each performance is 3.1%, and 3.6%. As for the precision performance value, both are almost the same, but the YOLOv5 is just slightly greater than the YOLOv8 model with a difference of 0.4%.