Computer Vision – YOLOv5 (April 2023)

Jorge A. Mejia ,Jorgemejia@cpp.edu , Knikolas Wooden, kkwooden@cpp.edu

Engineering department,California State Polytechnic University, Pomona

***Abstract*—This paper explores the performance of YOLOv5 across different computer architectures. The YOLO algorithm is an object detection framework that uses a single convolutional network and YOLOv5 is the latest object detection model developed by ultralytics. We tested YOLOv5 by doing object detection on live video, pictures, and pre-recorded videos from YouTube.**

***Keywords*—** Performance,clock rate , architecture

I. Introduction

Different architectures have different performances. A system's clock rate , number of instructions used , and Cycles per instruction(CPI) all have an effect on the performance as can be seen in equation 1.

$\frac{1}{Performance}$=execution time=

$$\frac{CPUClock\ cycles}{Clock\ rate} = \frac{Cpi * \#\ of\ instructions}{clock\ rate} \quad (1)$$

"Image classification involves assigning a class label to an image, whereas object localization involves drawing a bounding box around one or more objects in an image. Object detection is more challenging and combines these two tasks and draws a bounding box around each object of interest in the image and assigns them a class label. Together, all of these problems are referred to as object recognition.[3]" Object detection has some real life applications."Self-driving cars depend on object detection to recognize pedestrians, traffic signs, other vehicles, and more. For example, Tesla's Autopilot AI heavily utilizes object detection to perceive environmental and surrounding threats, such as oncoming vehicles or obstacles.[1]" Strategically placed people counting systems throughout multiple retail stores are used to gather information about how customers spend their time and customer footfall. AI-based customer analysis to detect and track customers with cameras helps to gain an understanding of customer interaction and customer experience, optimize the store layout, and make operations more efficient. A popular use case is the detection of queues to reduce waiting time in retail stores.[1]" "Object detection has allowed for many breakthroughs in the medical community. Because medical diagnostics rely heavily on the study of images, scans, and photographs, object detection involving CT and MRI scans has become extremely useful for diagnosing diseases, for example, with ML algorithms for tumor detection.[1]"

II. Objective

The main objective is to study the performance of YOLOv5 across different computer architectures using different methods.

A. YOLO

YOLO is an acronym for "You Only Look Once" and it was named that because it is a real time object detection algorithm that processes images at a very high speed. This state of the art, real time object detection algorithm was introduced in 2015 by Joseph Redmon, Santos Divvala, Ross Girshick, and Ali Farhadi in their famous research paper "You Only Look Once: Unified, Real-Time Object Detection".

III. Architecture description

We used two architectures to compare for this project. First one which we will call "architecture A" was an AMD laptop with a clock rate of 2.20 GHz. The second architecture, Architecture B, was a intel laptop with a clock rate of 2.50 GHz

IV. YOLO run 1

For the first test we tested how yolov5 and the two different architectures handle many items together

Figure 1 Source: Adapted from [2] base image

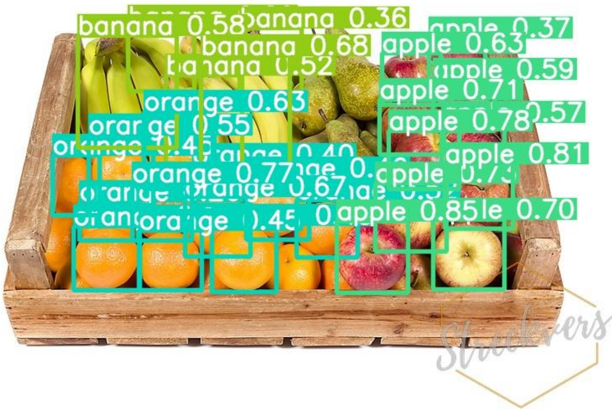

Figure 2 Architecture A

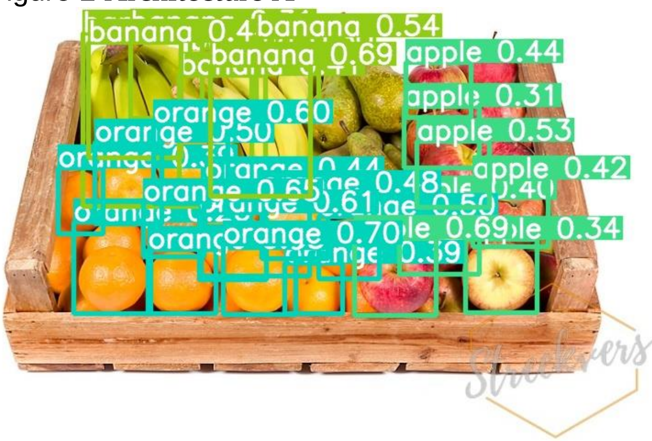

figure 3 Architecture B

Figure 1 was run through the two different architectures. Neither of the architecture could detect the pears. Architecture A had detected all the apples and gave a higher probability to all the apples. Some of the oranges are blocked from clear view so it is understandable how both architectures did not detect all the visible oranges. Architecture B detected more oranges. Architecture B seems to generally have higher probabilities for the oranges. Like the oranges , some of the bananas are obscured from view.

Neither architecture detected all the bananas. Architecture B detected more bananas and the probabilities for Architecture B also. For this image the execution time of Architecture A was 445.1 milliseconds. which according to equation 1 means 979,220,000 clock cycles were run.Architecture B had an execution time of 453.6 milliseconds. which according to equation 1 means 1,134,000,000 clock cycles were run.

The clock cycle tells us Architecture B needs to do more work to do the same thing as Architecture A. Architecture A seems to have the better performance but when including architecture A’s and architecture B’s 23.8 millisecond idle time and 2 millisecond idle time respectively, Architecture B is performing better.

V. YOLO RUN 2

For the next run we performed a test to compare how well the two different architectures recognize images under a blur.

Figure 4 Source: Adapted from [4] base image

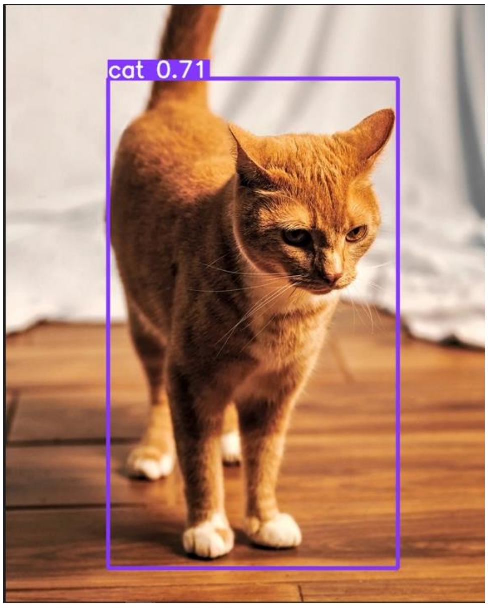

Figure 5 0% blur

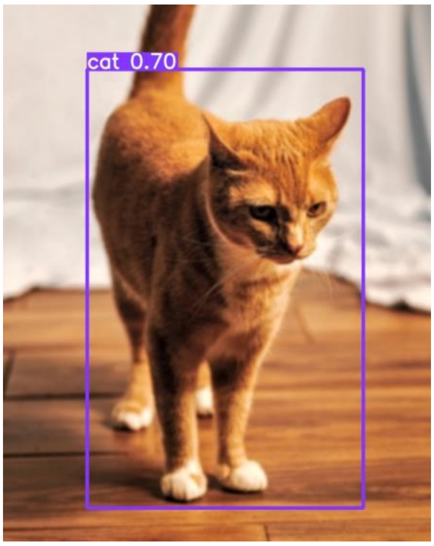

Figure 6 5% blur

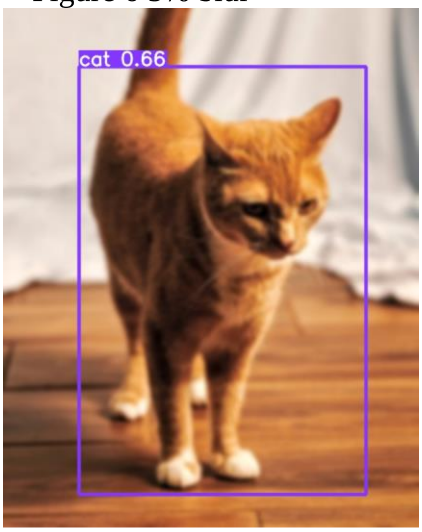

Figure 7 10% blur

Figure 8 15% blur

Figure 6 ,7, and 8 represent both the architecture as both architectures had the same percentages for all the cat images. We tested the blur at 0, 5,10,15, and 25 percent. Both architectures had no detection at 25 blur so the results were not used. Architecture A’s execution time for 0,5,10, and 15 were 380 milliseconds, 514 milliseconds, 515.6 milliseconds , and 368 milliseconds respectively with a idle time of 2 milliseconds, 3 milliseconds, 2 milliseconds, and 2 milliseconds respectively. Architecture B’s execution time for 0,5,10, and 15 were 681.9 milliseconds, 654.9 milliseconds, 1014.6 milliseconds , and milliseconds respectively with a idle time of 5.4 milliseconds, milliseconds,0 milliseconds, and 6.5 milliseconds respectively. For this image with one object to detect,Architecture A consistently had better performance regardless if idle time is considered.

VI. CONCLUSIONS

Seems like architecture A has better performance for smaller amount of objects but architecture A 's delay increases with multiple objects so much so that after a certain amount of objects architecture B performs better .

VII. REFERENCES

[1] G. Boesch, “Object Detection in 2021: The Definitive Guide,” viso.ai,Jul.09, 2021. https://viso.ai/deep-learning/object-detection/

[2] “Schoolfruit - Vers streekfruit op school,” Streekvers. https://www.streekvers.nl/product/schoolfruitpakket-gezonde-kleintjes/ (accessed May 03, 2023).

[3] Jason Brownlee, “A Gentle Introduction to Object Recognition With Deep Learning,” Machine Learning Mastery, Jul. 05, 2019. https://machinelearningmastery.com/object-recognition-with-deep-learning/

[4] S. Zhang, “How much would you pay to Save Your Cat's Life?,” *The Atlantic*, 16-Dec-2022. [Online]. Available: https://www.theatlantic.com/magazine/archive/2022/12/paying-for-pet-critical-care-cost-health-insurance/671896/. [Accessed: 06-May-2023].