# Comparative Studies for YOLOv4 Target Tracking Algorithm Performance

**Aurel Constantinescu, Rodica Constantinescu, Mihai Constantinescu**

Faculty of Electronics, Telecommunications and Information Technology, Polytechnic University of Bucharest, Bd. Iuliu Maniu, Nr. 1-3, Sector 6, Romania

## Introduction

This paper presents a detailed experimental study on operability of a target tracking system based on YOLOv4 for target detection. The authors propose that the results of each experimental complex scenario to be consolidated in 3D graph, named decision space, one for each environmental context, in which each point is the result of system efficiency reported to a specific pair of values of the mentioned algorithm parameters.

The results can be used as a method of pre calibrating the target tracking system prior of effectively using it for its main purpose.

## YOLOv4 architecture

YOLOv4 module detector is composed of two parts, a backbone which is pre-trained on ImageNet and a head which is used to predict classes and bounding boxes of objects. It is a one-stage object detector.

YOLOv4 consists of:

- Backbone: CSPDarknet53
- Neck: SPP, PAN
- Head: YOLOv3

The main purpose of CSPDarknet53 is to enable the architecture to achieve a richer gradient combination while reducing the amount of computation, achieved by partitioning feature map of the base layer into two parts and then merging them through a proposed cross-stage hierarchy.

The SPP module was originated from Spatial Pyramid Matching (SPM), and SPMs original method was to split feature map into several *d x d* equal blocks, where *d* can be {1, 2, 3, …}, thus forming spatial pyramid, and then extracting bag-of-word features.

## Proposed approach

Proposed approach related to optimum operability of a target tracking system based on YOLOv4 for target detection, emphasis the importance of detection phase on the performance of the overall system and how that performance can be evaluated and how the optimal zones, named operational zones, can be identified based using decision spaces.

This work aimed to determine multi-dimensional regions where different parameters from the algorithm and related to the real-life context of monitoring target have influence on the overall target recognition performance in different specific real time situations.

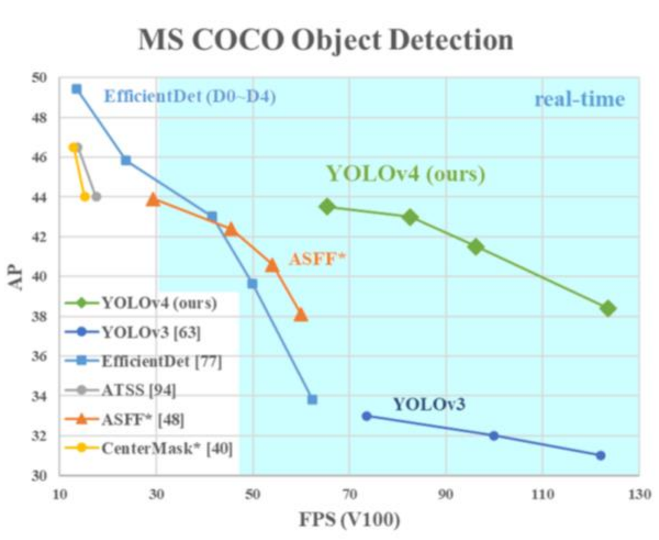
Given the high number of combinations of the above-mentioned parameters, each of them influencing in an inter-dependent mode the overall performance of the system, experimental studies of complex scenarios are very helpful to determine the combined effect on such parameters.

In this work, the experiment' results are consolidated in multidimensional graphs, named decision spaces. These decision spaces present in aggregated form the performance obtained in a target tracking scenario, related on a specific situation. The results can be used as a method of pre calibrating the target tracking system prior of effectively using it for its main purpose.

Based on decision spaces, both optimal parameter set, and limited performance "zone" can be identified, providing to system user a detailed view of system limitations and performances, according with real life environment within the system will perform.

## YOLOv4 comparative performance

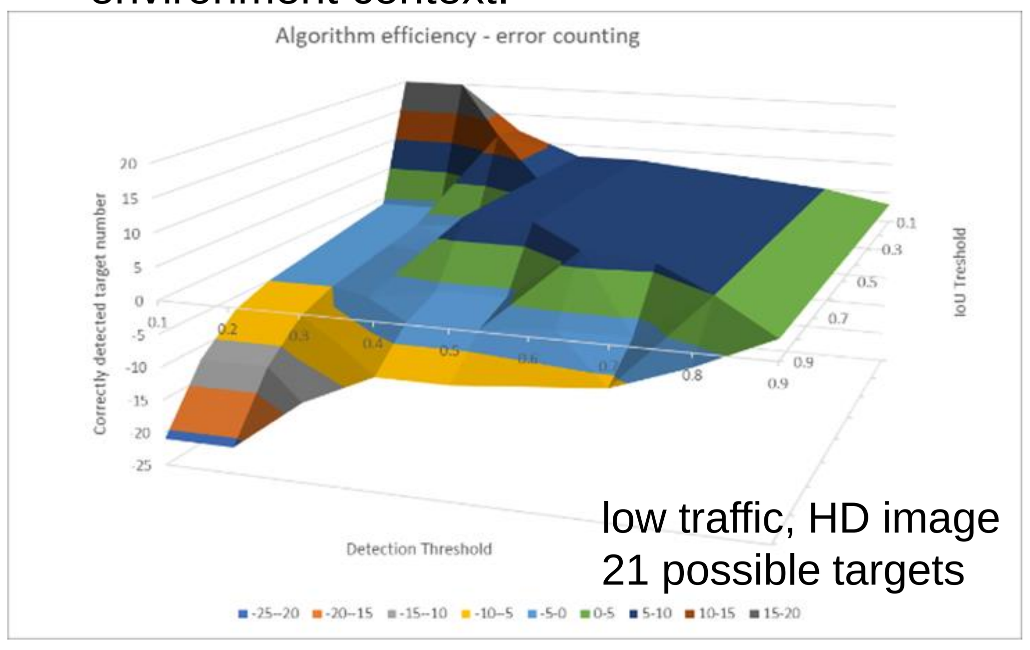
The optimal balance among the input network resolution, the convolutional layer number, the parameter number (filter size2 * filters * channel / groups), and the number of layer outputs (filters) are considered, together with comparative results with the other most popular algorithms.

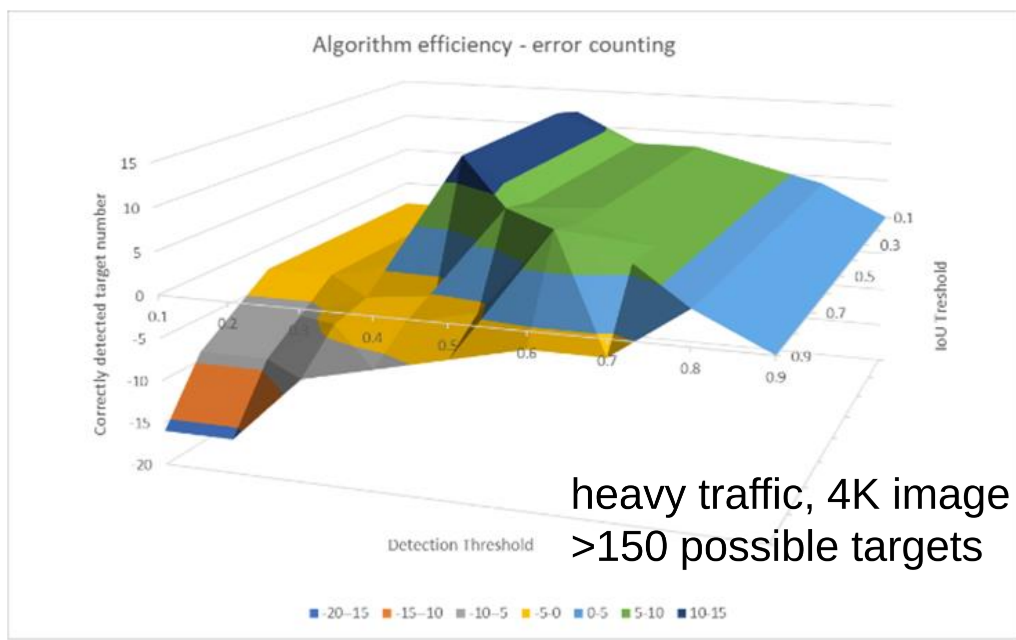



## Results – system performance

Due paper size limitations, only two types of scenarios are presented: low traffic, HD image; and full HD image, heavy traffic, 4K image.

For further development, a special attention is paid to system functionality, so, in future, with proper modification, several errors such as some wrong identification window could be corrected, and the system performance could be maximized according with the environment context.




## Experimental Studies Context

That research is focused on about how many targets could be detected in a frame/image, and about the range of relative size of targets toward frame/image dimensions that assure the target detection using YOLOv4.

The detection performance depends on video quality, the observation angle and YOLOv4 internal parameters: detection threshold (noted detTh) and Intersection over Union (IoU) threshold (noted oTh) combined effects on algorithm detection performance. It is important to observe that algorithm/system performance is not computed related to YOLO accuracy, but to wrong detection avoidance capability and the total number of correct detections in a specific environment and system configuration.

## Conclusions

The study's results highlight a high performance for images with low number of targets to a very low performance for high traffic, 4K image quality. An HD image, with few targets, it is processed using several anchors, well placed due the low number of details provided by the image. As far as the possible target numbers are increasing, as is the case for medium traffic scenarios, the anchor's approach and the detections results are not so accurate, and the overall system performance decrease to half compared with the previous case.

Authors of this paper have only in mind to identify the performance of a target tracking algorithm based on YOLOv4 for detection on different real environments, and to open new domains of research for future system optimization and new object detection algorithm capabilities.

## Literature cited

- A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection", arXiv preprint arXiv:2004.10934.
- C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696.
- M. Constantinescu and E. Borcoci, "Optimization method for smart antenna use in WiMAX vehicular communications," 2012 9th International Conference on Communications (COMM), 2012, pp. 263-266, doi: 10.1109/ICComm.2012.6262555.

ATOM-N 2022
11th edition