**FlySight project – Object detection**

# Introduction

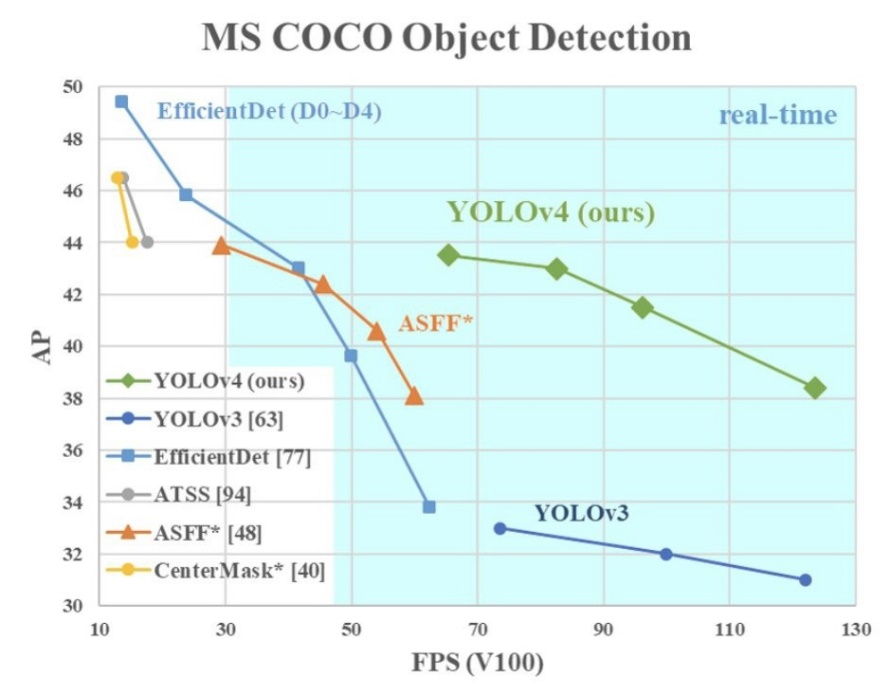
Search And Rescue (SAR) operations in non-urban areas (mountain, forest, canyon, water body…) usually requires a large amount of human personnel and when using drones, even an operator able to focus the entire time of the operation could potentially fail to identify a body in an unusual situation. Automated object detection could thus help save lives in such situations.

In this paper, we will be comparing the performance of 2 different object detectors : YOLOv5 and SSD. We will first briefly present the structure of the 2 nets used afterwards. Then we will introduce our database before showing the explorative data analysis done on this set. Next we will explain the metrics used to compare the nets and the different trains. Eventually, we will showcase the performance of the nets and the influence of the train parameters for SAR operations.

# Presentation of the nets

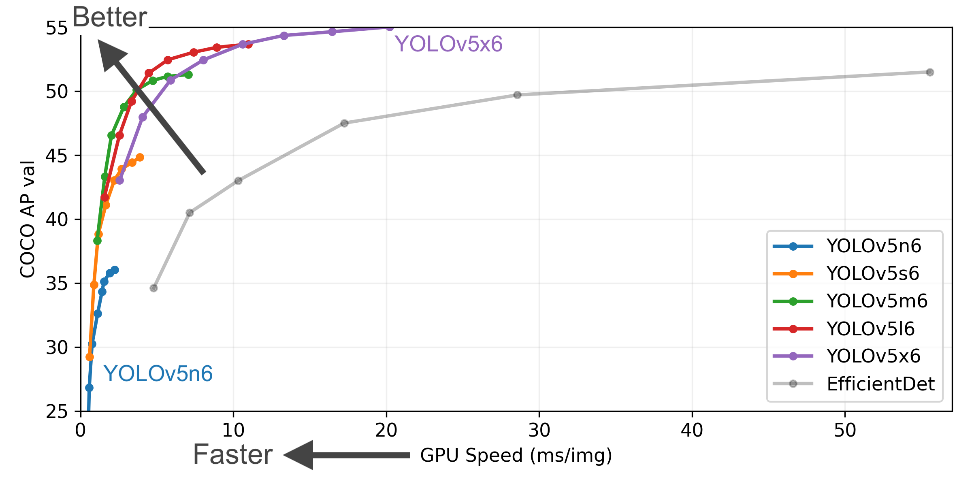
## YOLOv5

YOLOv5 is a state-of-the-art detector, it is based on the PyTorch framework. This detector is a You Only Look Once type of detector meaning the full images are processed by the neural network instead of applying the network multiple times to the image at different locations and scales. This network has one of the best speed on the market (*Figure 1*) with a relatively good precision.



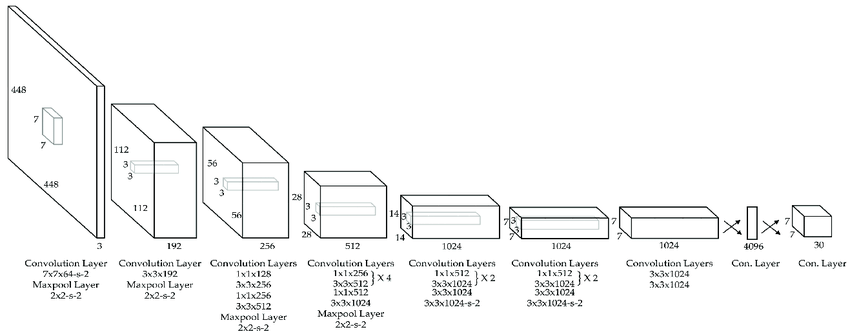
*Figure 1: Comparison of YOLOv4 and other state-of-the-art detectors*

Since detection speed is a key factor is SAR operations, we will be testing YOLOv5 as the best compromise between speed and accuracy amongst the YOLO versions. The main difference between YOLOv5 and previous YOLO versions is the translation of the Darknet framework to PyTorch framework. PyTorch framework speeds up the training and the inference using 16 bits floating point precision instead of the 32 bits used in the Darknet framework. To keep the training time reasonable, we will be using the YOLOv5n network which is the smallest and the fastest of all YOLOv5 nets (*Figure 2*).



*Figure 2: YOLOv5 nets comparison*

YOLOv5 uses *New CSP-Darknet53* as the backbone, *SPPF* and *New CSP-PAN* as the neck and *YOLOv3 Head* as the head. The architecture is a deep convolutional network as shown on *Figure 3*.



*Figure 3: High level diagram of YOLO architecture*

## SSD

TODO

# The dataset

The data we will be using for training and testing is the SAR dataset (SARD) built by FlySight. The images in this set present aerial view of persons in typical SAR situations: walking, running, sitting, standing or lying down. The database is made of 1920x1080 images and contains 1980 annotated frames (typical SAR situation on *Figure 4*).



*Figure 4: Image “gss.1” from the SARD database*

In the SARD database, the bounding boxes dimensions go from 7px for width and 8px for the heigh to 353px for the biggest width and 337px for the biggest heights. The area of bounding boxes range from 7x12px for the smallest to 322x231px for the biggest while the average area is 47x58px (according to FlySight article). In terms of data, there are 6532 person objects spread unevenly in terms of size.

In the original database that was given to us by FlySight, the labels were stored using XML files in PASCAL VOC format.

The SARD database is unique in its kind because it is specifically adapted for SAR operations. Other drone-made databases on the market (VisDrone, Okutama-action, UAV123, CARPK dataset, ERA dataset…) achieve good results in detection of person when used for training neural networks. However all of them uses image from urban areas (*Figure 5*), thus reducing the efficiency of the detector when used in non-urban areas.



*Figure 5: Inference on a VisDrone dataset image*

# Explorative Data Analysis

## Format the labels

# Metrics

# Results

# Conclusion