Object Detection for Rescue Operations by High-altitude Infrared Thermal Imaging Collected by Unmanned Aerial Vehicles

Andrii Polukhin[1], Yuri Gordienko[1], Gert Jervan[2], and Sergii Stirenko[1]

[1] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[2] Tallinn University of Technology, Tallinn, Estonia

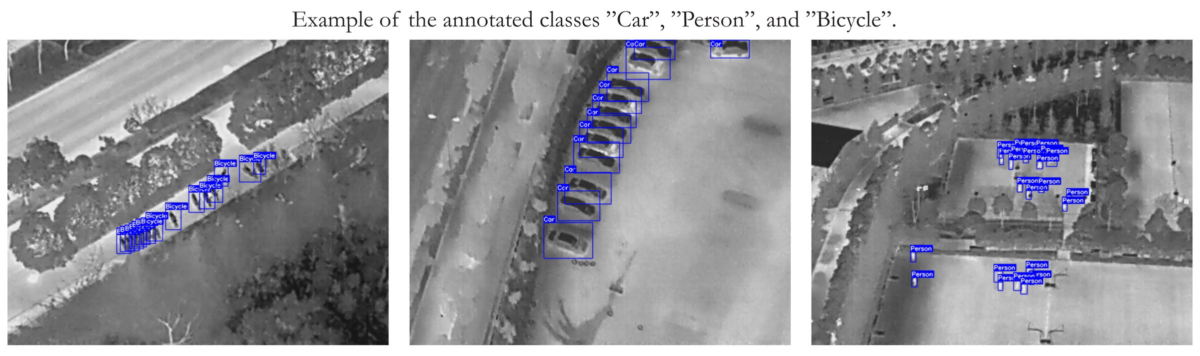
Abstract

The analysis of the object detection deep learning model YOLOv5, which was trained on High-altitude Infrared Thermal (HIT) imaging, captured by Unmanned Aerial Vehicles (UAV) is presented. The performance of the several architectures of the YOLOv5 model, specifically 'n', 's', 'm', 'l', and 'x', that were trained with the same hyperparameters and data is analyzed. The dependence of some characteristics, like average precision, inference time, and latency time, on different sizes of deep learning models, is investigated and compared for infrared HIT-UAV and standard COCO datasets. The results show that degradation of average precision with the model size is much lower for the HIT-UAV dataset than for the COCO dataset which can be explained that a significant amount of unnecessary information is removed from infrared thermal pictures ("pseudo segmentation"), facilitating better object detection. According to the findings, the significance and value of the research consist in comparing the performance of the various models on the datasets COCO and HIT-UAV, infrared photos are more effective at capturing the real-world characteristics needed to conduct better object detection.

Introduction

Unmanned aerial vehicles (UAVs) are frequently used in many different industries, such as emergency management or environment monitoring. Infrared thermal imaging has been identified as a promising method for enhancing object detection in adverse weather conditions and challenging environments. The significance and value of our research are to provide insights into the potential use of infrared thermal imaging on UAVs for object detection in challenging weather conditions, particularly in rescue operations. Special attention will be paid to the investigation of the dependence of some characteristics (average precision, inference time, and latency time) on different sizes of deep learning models, comparing infrared and standard datasets.

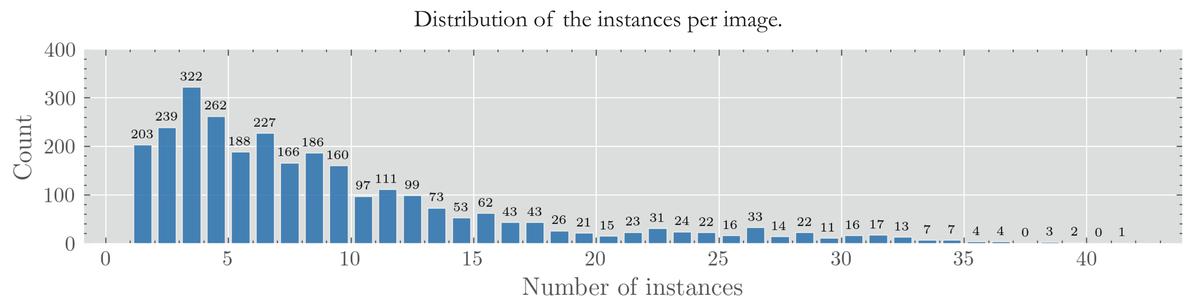
Example of the annotated classes "Car", "Person", and "Bicycle".



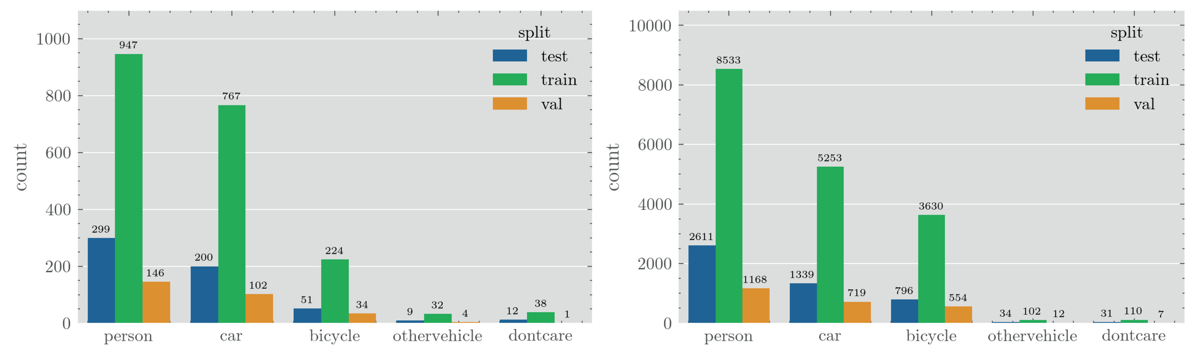
Methodology

We conducted an analysis of the distribution of the five annotated object categories across instances and images of the HIT-UAV dataset, which includes five categories: Person, Car, Bicycle, Other Vehicle, and DontCare. For subsequent experiments, all five categories were used for training, but only the top three categories were used for evaluation. This is because there is a shortage of training data for the OtherVehicle and DontCare categories, which limited the model's ability to learn them effectively.

Distribution of the instances per image.

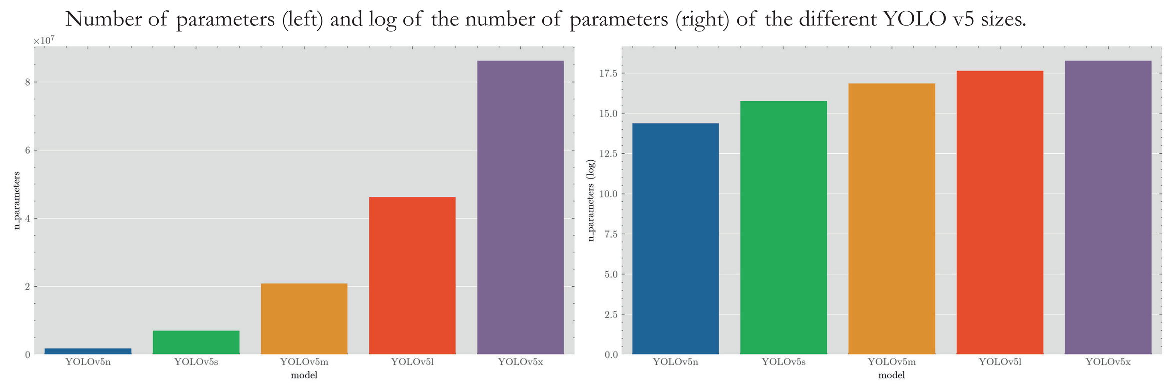


Distribution of the categories across the instances and images.



We utilized the standard and freely available YOLOv5 single-stage object detector. YOLOv5 offers five different sizes of architecture, ranging from the extra small (nano) size model denoted by 'n', to the small ('s'), medium ('m'), large ('l'), and extra large ('x') models.

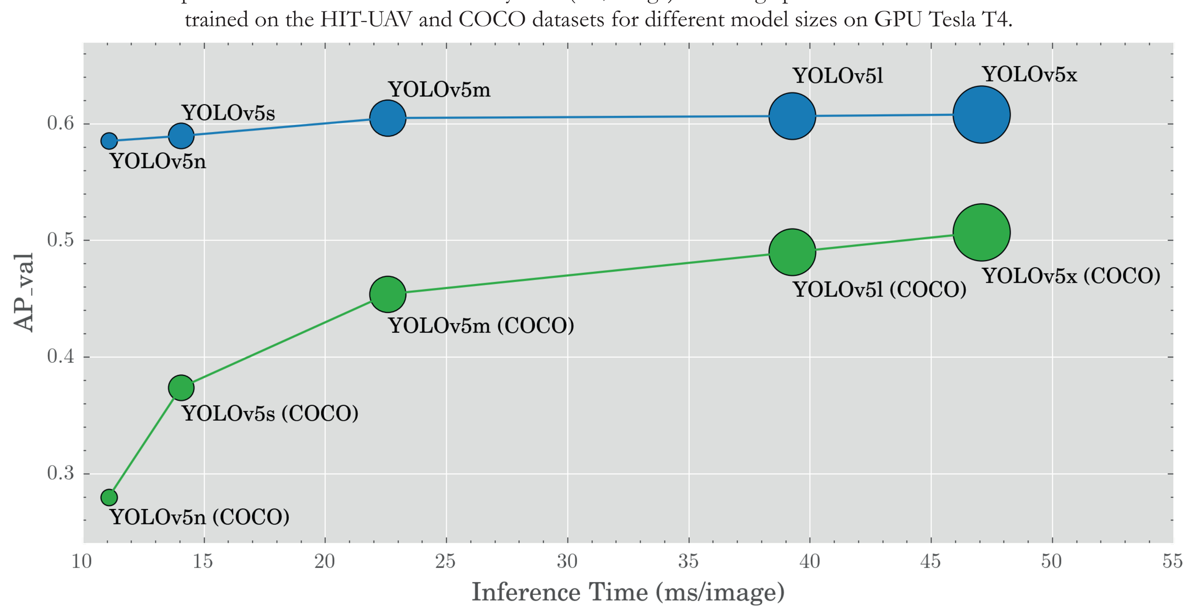
Number of parameters (left) and log of the number of parameters (right) of the different YOLO v5 sizes.



In addition, we compare our results with YOLOv5 models trained on the COCO 2017 dataset. The models were trained for 300 epochs using A100 GPUs at an image size of 640. Our training setup closely follows this configuration. These models are used to evaluate the performance of real and infrared thermal imaging.

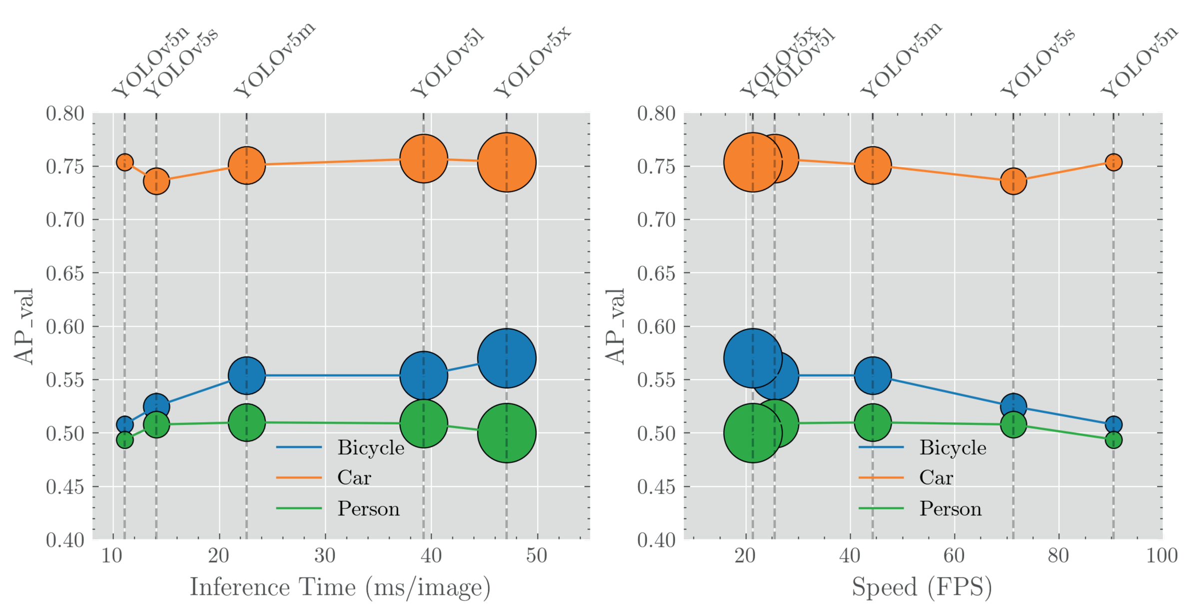
Results

Comparison of the inference and latency time (ms/image) vs average precision of YOLOv5 models trained on the HIT-UAV and COCO datasets for different model sizes on GPU Tesla T4.



YOLOv5 performs better on the HIT-UAV dataset than on COCO, indicating the advantages of using IR aerial imagery over RGB imagery for object detection tasks for certain classes. It indicates the following observations: (1) A significant amount of unnecessary information is removed from infrared thermal pictures, providing a "pseudo segmentation" effect and facilitating object detection. Due to the clearly visible properties of the objects in infrared thermal pictures, the typical detection model may achieve great recognition performance with few images; (2) In the infrared thermal aerial photos, we can see that the model does great of capturing large objects. Small objects, including persons, might easily be mistaken for infrared sensor noise, which raises the false positive rate.

Comparison of the inference time (ms/image) and speed (FPS) vs average precision on different object categories of different model sizes on GPU Tesla T4 for HIT-UAV.



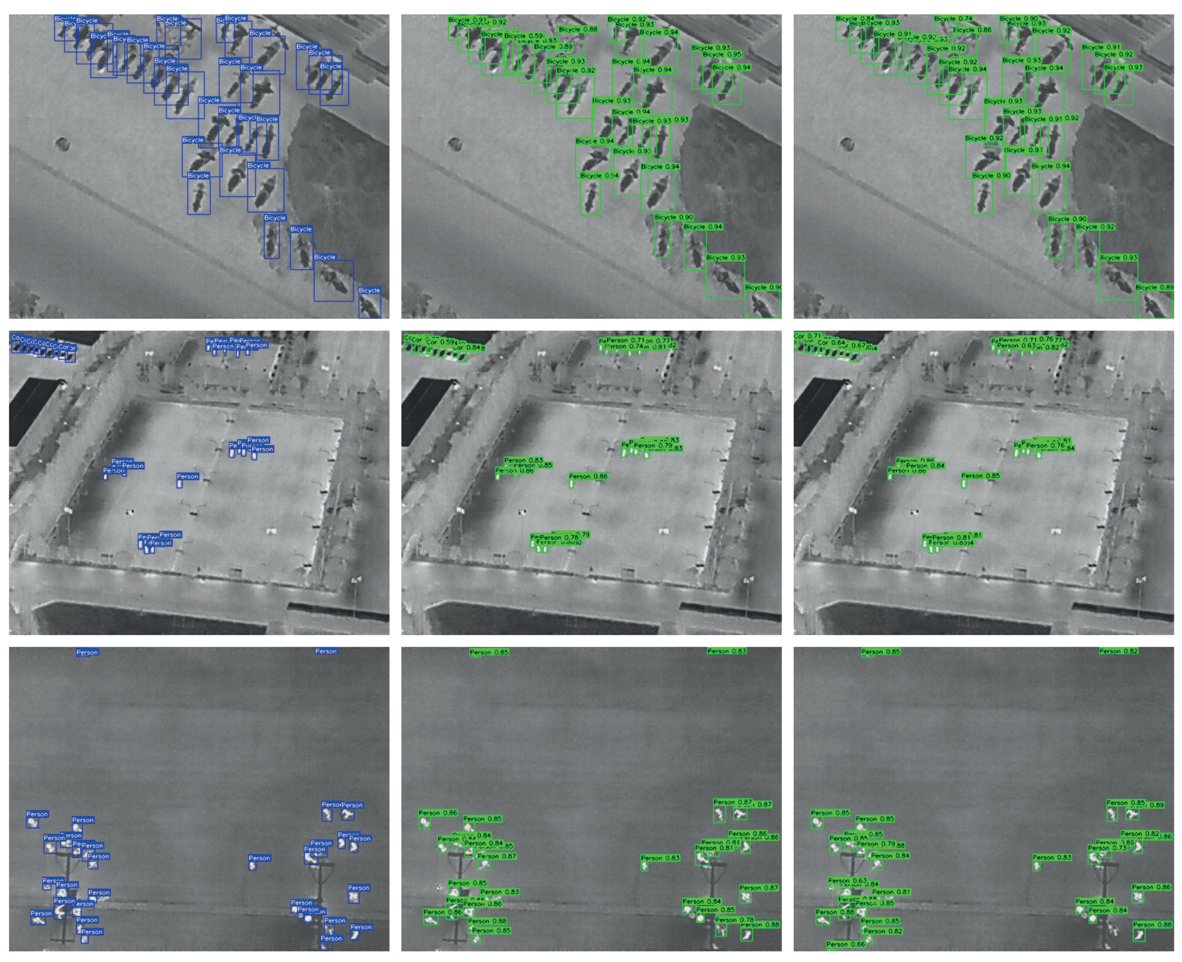
The evaluation for the YOLO v5.

Model	Dataset	Person	Car	Bicycle	FPS	AP
YOLO v5 (n)	HIT-UAV	49.4%	75.4%	50.8%	90	58.5%
YOLO v5 (s)	HIT-UAV	50.8%	73.6%	52.5%	71	58.9%
YOLO v5 (m)	HIT-UAV	**51.0%**	75.1%	55.4%	44	60.5%
YOLO v5 (l)	HIT-UAV	50.9%	**75.7%**	55.4%	25	60.6%
YOLO v5 (x)	HIT-UAV	50.0%	75.4%	**57.0%**	21	**60.8%**
YOLO v5 (n)	COCO	-	-	-	90	28.0%
YOLO v5 (s)	COCO	-	-	-	71	37.4%
YOLO v5 (m)	COCO	-	-	-	44	45.4%
YOLO v5 (l)	COCO	-	-	-	25	49.0%
YOLO v5 (x)	COCO	-	-	-	21	50.7%

Conclusion

Our findings reveal that the HIT-UAV dataset shows a less degradation of average precision with decreasing model size compared to the COCO dataset. This can be attributed to the removal of extraneous information from infrared thermal images, resulting in better object detection performance for certain classes, such as persons who emit heat or vehicles with metal signatures highlighted in infrared images.

Sample results of the YOLO v5. Left is the original annotations, the middle is the smallest YOLO v5 (n), and the right is the largest YOLO v5 (x).



Moreover, our study shows that infrared thermal images significantly enhance object detection capabilities by filtering out unnecessary information and improving the recognition of certain classes compared to visual light images. These benefits increase the feasibility of autonomous object detection using UAVs in crucial nighttime activities, such as surveillance, person search, and rescue.

This publication is supported by: NATO OTAN The NATO Science for Peace and Security Programme (grant G6032)