PaparazzoDrone

Grant Tannert, Tim Mandzyuk
University of Washington
Seattle, 98105
guranto@uw.edu, timmandz@uw.edu

Abstract

The potential for drones as autonomous cameramen is endless. In this report, we explore the limitations and challenges of using a drone for human detection, tracking, and drone stability. Specifically, we use the DJI Tello drone, as it's lightweight, easy to use, and affordable. We explore the challenges posed by the drone's camera quality, processor, and instability in flight. We briefly explored HAAR Cascade for object detection before switching to and focusing on YOLOv5 as our main model. Our findings show that the combination of our drone and object detection model provided a few limitations. The object detection worked very well and the image quality was great, but the stability of our drone was lacking and processing was slow. By looking closer at these observations, our report aims to give insight on the feasibility and potential enhancements to improve the performance of the DJI Tello drone as an autonomous cameraman.

1 Introduction

In recent years, drones have emerged as a popular technology with many use cases across all industries. Today we see them used for package delivery, search and rescue operations, detecting wildfires, etc. A new field we see being revolutionized by drones is the field of being a cameraman. Imagine a drone seamlessly flying by a track athlete running or recording a surfer shredding some gnarly waves. Traditional cameramen would not be able to capture these moments, or if they do, they require rigorous setups. The possibilities seem endless with such a drone, but getting there with precision is no easy feat.

To do this, we would need to consider various tracking techniques for capturing the motion of individuals in real-time. The issue would not only be with tracking, but also detection and stability. An even greater challenge posed here is doing this in dynamic environments. Despite all the challenges, the integration of drones as autonomous cameramen offers promising solutions to challenging problems.

2 Research Question

One notable drone model that stood out was the DJI Tello, which is known for its small size and ease of use. The challenge here is leveraging the Tello drone for object detection, tracking, and stability. For our project, we aim to answer the research question of, what are the limitations and challenges of using a DJI Tello drone for human detection, tracking, and stability? The drone faces a few main limitations starting with camera quality. We want to explore the impact of the DJI Tello's camera on image resolution, as it might be lower resolution than high-end drones. We also need to consider the CPU as a potential limitation, as it's lightweight and less powerful. It uses the Intel Movidius Myriad 2 VPU. This may affect the ability for real-time object detection and tracking efficiently. Not just this, but the stability of the Tello drone is another concern. We wonder if instability issues will

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

cause blurry photos in the image capturing process. By investigating the camera quality, processing limitations, and drone stability, we hope to understand the constraints of the drone and their effects on the quality of tracking humans. We hope our findings will reveal how feasible it is to make a quality autonomous cameraman, and shed light on how reliable the DJI Tello drone is.

3 Related Work

Object detection is widely and commonly used in remote sensing fields to do tasks such as autonomous navigation, face recognition, object detection, safety systems, and etc. Traditional object detection techniques such as HOD, Haar Cascades, and SIFT suffer from limited representation power because features are handcrafted and may not capture the complexity of data, and limited scalability because feature extraction is complex and may cause feature dimension explosion resulting in computational and memory bottleneck. Deep learning techniques for object detection have shown to produce more accurate and faster results, and thus resulting in the advent of deep learning and neural networks for object detection. There are two major categories of deep learning techniques for object detection: 1. two-stage networks, 2. one-stage networks. 1. A two-stage network, such as RCNN, Fast RCNN, Faster RCNN, generates a region proposal in the first stage, then classifies and locates candidate regions in the second stage. 2. A one-stage network, such as SSD and YOLO, directly classifies and generates class probability and coordinate locations. Two-stage networks typically have higher detection accuracy while one-stage networks are faster and more lightweight. In our application, we address issues such as low-computational resources and fast real-time detection, thus using a one-stage network is a better fit. Hence, in this paper we explore the YOLOv5 algorithm for our object detection model.

M. Liu et al.[1] used the YOLOv3 model for UAV small object detection where they improved the model by adding convolution operations at an early layer to enrich spatial information and enlarging the receptive field. Y. Yang[2] used the YOLOv5 model for UAV object detection where they added an unsampling operation to the neck of the network which generates a feature map for collecting small target features and they added an image segmentation layer before the detection layer to reduce the feature loss of the image in the downsampling and improves the detection effect. L. Zhu[3] modifies the YOLO model by including a new backbone Darknet59; added a new complex feature aggregation module MSPP-FPN that incorporated one spatial pyramid pooling and three atrous spatial pyramid pooling modules; and used a Generalized Intersection of Union loss function. Y. Li[4] used the YOLOv8 model for UAV object detection where they introduced a Bi-PAN-FPN to improve the neck part in YOLOv8. By fully considering and reusing multiscale features, a more advanced and complete feature fusion process is achieved while maintaining parameter costs. They also used a WiseIoU loss function for bounding box regression loss, combined with a dynamic nonmonotonic focusing mechanism, and the quality of anchor boxes is evaluated by using "outlier" so that the detector takes into account different quality anchor boxes to improve the overall performance of the detection task.

4 Methodology

YOLOv5 offers different pre-trained models of varying sizes, labeled YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, which are in increasing order relative to the model size. In our application, the DJI Tello has a lightweight processor and we need real-time detection, thus we chose to use the smallest model, YOLOv5s. We will first discuss our people detection approach, then proceed to our algorithm for following the person.

First, we gather images from the video frames fed from the DJI Tellos camera. The video frame rate is 30fps with each frame being 1280 x 720p HD. Then, we feed the image into the pre-trained YOLOv5s model which outputs a pandas dataframe. From the pandas dataframe, we extract the rows where the object type "people" is detected and ignore other rows. For each "people" row, we save the information about the position of the objects and the confidence of the classification. We then set a hyperparameter for the confidence threshold, which we set as 0.6 for our implementation. For any "people" with confidence over the threshold, we surround them with bounding boxes and highlight the center with a filled circle. Then, we select the person to follow which will be the person with the highest confidence rating. If there are ties, then we select the first person detected by the model.

Now we discuss our algorithm for following the selected person. After the person has been selected, we try to keep the person's center at the center of the DJI Tello's camera view and keep the person within 200-300 cm from the drone. In order to do this we will use Python's simple-pid module to incorporate PID controls in our DJI Tello's movement. First, we take the difference of the x,y position from the person's center and the camera's center, this will be the error for PID. Then, we calculate the amount of yaw angle to turn and the amount of up/down to move and send these commands to the DJI Tello. In order to determine the person's distance from the drone, we tested different distances from the drone and the bounding box's area produced by the distances. After several tests, we were able to find that the bounding box area of [15000px, 20000px] roughly estimates a distance of 200-300 cm from the drone. Then, if the bounding box of the person is too small, we send a forward movement to the DJI Tello. If the bounding box is too large, we send a backward movement to the DJI Tello. Repeat this process of centering and forward/backward, until the desired state is reached.

5 Experiments

As mentioned in the introduction, we are using a DJI Tello, which contains the following specification. The DJI Tello is a small, lightweight(80g) quadcopter with max speed 8m/s. It contains a camera with 720p, HD and 30fps. Its processor is the Intel Movidius Myriad 2 VPU which is a lightweight processor. In our experiment, we wanted to focus on common problems faced in object detection with UAVs. These problems are: object detection on small objects; object detection on direction diversity (recognizing the same object facing different directions); object detection on scale diversity (recognizing the same object scaled differently); real-time object detection, and image degradation (lighting, weather, etc. affecting image quality). Thus, in our experiment we focus on three scenarios, how well object detection under (1) lighting, (2) occlusion, (3) scale. We used the confidence measure from the output of the YOLOv5 model in order to determine the performance under these conditions. We took these experiments, in a closed room with no weather effects and kept other variables controlled when focusing on one variable.

Table 1: Confidence under different lighting.

Distance	Bright	Semi-lit	Dark
Far	0.90658	0.86560	0
Close	0.93039	0.78093	0.38181

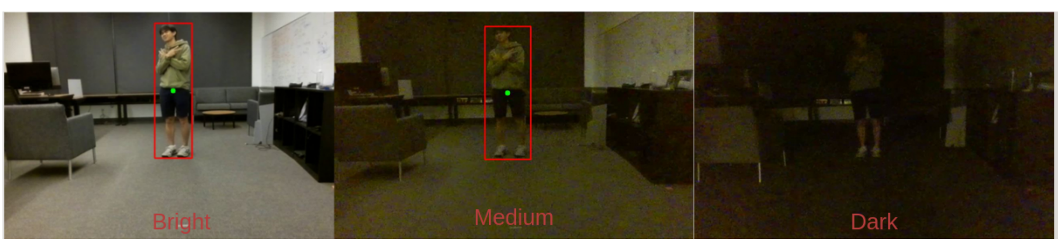

Figure 1: Drone's camera with bounding boxes under different lighting.

In Table. 1 and Fig. 1, we can see that the YOLOv5 model is still able to detect the person(confidence above threshold) within a bright and semi-lit room, however isn't able to do so in a dark room.

Table 2: Confidence under various occlusion and scale.

None	Some	Some More	Most	Big	Small
0.93039	0.88969	0.88171	0.45313	0.69933	0.68316

Figure 2: Drone's camera with bounding boxes under various occlusion.

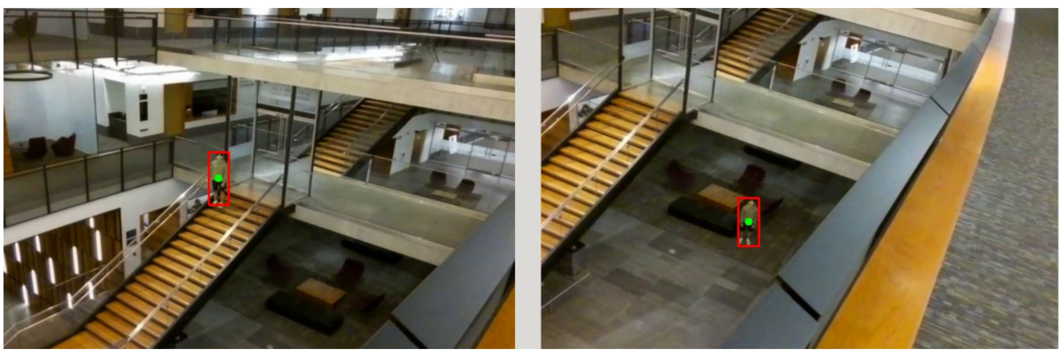

Figure 3: Drone's camera with bounding boxes under different scales.

In Table. 2 and Fig. 2, we can see that the YOLOv5 model is able to detect the person(confidence above threshold) when the object was somewhat occluded, but fails to do so when the only the person's face was showing. Additionally, we can observe from Table. 2 and Fig. 3 that the YOLOv5 works well in detecting the object when it's scaled in different size as it still has a confidence above the threshold.

From our results, we can determine that using the YOLOv5 model for object detection produces amazing results as it's able to keep track of the person under various settings. However, we can see that there are still limitations for example if the drone were to go under a tunnel it may lose track of the person, or if the person was a really far and was the size of a head like in Fig 2., then it will not be able to find the person. The first limitation can be addressed by possibly adding spatial and temporal locality in object detection thus predicting the trajectory of the person. The second limitation may be addressed by possibly adding by adding convolution operations at an early layer like M. Liu et al.[1] to enrich spatial information and enlarging the receptive field.

6 Conclusion

In this paper we addressed that the YOLOv5 model worked well as an object detection model for the DJI Tello drone under ideal environment conditions (a closed room interview, or a building survey) and allowed for fast, consistent, real-time detection. However, it suffers when the environment undergoes harsh changes which may cause image quality to drop.

References

[1] M. Liu, X. Wang, A. Zhou, X. Fu, Y. Ma, and C. Piao, "Uav-yolo: Small object detection on unmanned aerial vehicle perspective," Sensor, vol. 20, no. 8, p. 2238, 2020.

[2] Y. Yang, "Drone-View Object Detection Based on the Improved YOLOv5," 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), Changchun, China, 2022, pp. 612-617, doi: 10.1109/EEBDA53927.2022.9744741.

[3] Zhu, Li Xiong, Jiahui Xiong, Feng Hu, Hanzheng Jiang, Zhengnan. (2023). YOLO-Drone:Airborne real-time detection of dense small objects from high-altitude perspective.

[4] Li Y, Fan Q, Huang H, Han Z, Gu Q. A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition. Drones. 2023; 7(5):304. https://doi.org/10.3390/drones7050304