DRONE-BASED HUMAN DETECTION FOR SEARCH AND RESCUE WITH OBJECT DETECTION AND SLICING AIDED HYPER INFERENCE

Clent Japhet A. Poledo

INTRODUCTION

Unmanned Aerial Vehicles (UAV) are becoming more and more popular for a multitude of commercial and civil applications such as smart agriculture (Tripicchio et al., 2015), delivery (Bevacqua et al., 2015), and search and rescue (SAR) operations (Alotaibi et al., 2019; Bevacqua et al., 2015; Scherer et al., 2015; Silvagni et al., 2017). Traditional search and rescue methods requires the rescue team to be in the point of interest as fast as possible because time is critical in this field – delay can potentially lead to human losses (Waharte and Trigoni, 2010). For example, during an earthquake, the survival rate of victims lowers down very quickly as time goes on. In 24 hours, the chance of a victim surviving is 74%, 22% in 72 hours, and 6% in 120 hours (Chiu et al., 2020). Using UAV is one approach as a supporting method for SAR operations as these are fast and agile. This leads to being able to survey areas of interest faster and detect the location of victims ahead of time. Therefore, in the context of SAR operations, the ability of UAV to detect the presence of a potential victim is highly important.

These vehicles are equipped with technologies such as sensors to gather information about its environment. The camera is one of the most important input devices that is used in SAR operations. Manually screening the images for detecting persons is slow and subject to human error, therefore, algorithms were developed to facilitate the analysis of the images (Božić-Štulić et al., 2019). The usage of object detection techniques is one of the strategies used for the detection of potential victims. Modern techniques for object detection uses deep learning approach which produces state-of-the-art results (Marušić et al., 2018).

There are recent studies that explored the capabilities of deep learning object detection algorithms for human detection in the context of SAR. In a study by Mishra et al. (2020), they trained deep learning object detection models that detect humans and their poses such as standing, running, waving, etc. Rizk et al. (2021) trained a YOLOv3 model for detecting humans at a low-altitude shot. Qingqing et al. (2020) analyzed the performance of YOLOv3 at varying height for detecting humans for SAR, specifically marine search. Božić-Štulić et al. (2019), Domozi et al. (2020), Dousai and Lončarić (2022) and Marušić et al. (2018) experimented on various object detection models using the HERIDAL dataset, a dataset that features a high altitude, high quality, and a well-labeled set of images simulating SAR situations in non-urban areas. However, the researcher found some gaps that need to be addressed.

From the studies mentioned previously, many studies compared the models in terms of mean Average Precision (mAP) and failing to document the recall metrics. While mAP is the standard metric when dealing with object detection, it should be worth noting that recall is very important in the context of SAR (Božić-Štulić et al., 2019; Domozi et al., 2020). Moreover, some studies used a dataset based on closer, low-altitude shots. While it might be good for surveillance, a bigger field of vision is more efficient for SAR. Lastly, a high-altitude drone perspective can be a problem in training. In this case, objects will look smaller and it will be more difficult to detect (Chen et al., 2022; Liu et al., 2021). On top of that, high-resolution images could produce errors during training if the GPU memory is not enough.

In this study, the researcher will train a model for human detection in high altitude drone shots with the motivation of being able to cover more ground as flying high enough can cover a larger area for a more efficient searching. The researcher will use the HERIDAL dataset which is a dataset crafted for the purpose of human detection in the context of SAR (Božić-Štulić et al., 2019). To address the issue regarding training and inferencing on high-resolution images, slicing aided hyper inference (SAHI) will be used. The SAHI library includes slicing of the dataset into smaller parts so that the dataset can be used for training. The main feature of SAHI is it allows for a more effective object detection on a high-resolution image, especially on relatively small objects, by inferencing on smaller slices of the image instead of simply resizing the image (Akyon et al., 2022). Additionally, the dataset slicing and inference slicing are independent which gives more flexibility to the study. The object detection model that will be used in this study will be YOLOv8, which is a real-time object detector. The recall of the model on different SAHI parameters will be properly documented and compared. mAP and inference time will also be noted as a reference for other related research.

**Objective of the Study**

This study applies SAHI to a deep learning object detection model to achieve high performance in detecting humans for SAR. Specifically, this study aims to:

1. implement YOLOv8 for human detection with the HERIDAL dataset with hyperparameter optimization;
2. test various SAHI parameters to improve the performance of the trained object detection model;
3. determine the effectiveness of the model with SAHI on the HERIDAL dataset by documenting and comparing inference time, recall, precision, and mAP to past studies utilizing the same dataset.

This study can help towards more effective SAR operations to events such as finding missing people. In addition, this study will also be able to contribute towards the sustainable development goals, specifically the third goal that aims towards ensuring the healthy lives and promoting the well-being of all people of all ages. Locally, the results of this research can be used by the local National Disaster Risk Reduction and Management Council (NDRRMC) and Harmonized Aerial Watch and Knowledge-Based Survey (HAWKS).

The scope of this study is only to explore the effectiveness of SAHI and YOLOv8 in detecting humans in the SAR context. Models that are not tested in this study are outside the scope of this research. Moreover, the dataset is composed of only drone shots from forested areas, therefore, the trained model cannot be generalized for human detection in general.

The rest of the paper is organized as follows. Chapter 2 is about the review of related literatures, which will be composed of studies related to the area of UAV, SAR, object detection, and its intersections. Lastly, Chapter 3 describes the methodology, which will show the framework on how the study will be conducted.

REVIEW OF LITERATURE

In this chapter, the concepts that are related to the study as well as related literatures will be thoroughly discussed. Specifically, this chapter will cover about search and rescue (SAR), the usage of UAVs for SAR operations, the use of object detection techniques to UAVs for human detection, the gaps from recent studies that uses human detection for SAR operations, the proposed method to address the gap, and thorough discussion on the models and optimizing techniques that will be used in the study. The topics that will be discussed in this chapter can help augment the reader’s knowledge of the concepts within the scope of this research – to train an object detection model for human detection from a high-altitude drone perspective and examine the effectiveness of slicing aided hyper inference on the performance of the model.

Search and Rescue

Search and rescue, or “SAR”, is an activity that is carried out around the world for centuries. According to the book “Fundamentals of Search and Rescue” by Cooper (2005), difficulties can be faced by humans when travelling to unknown terrains and assistance can be needed from other people with knowledge of the terrain as well as having appropriate equipment. Government-supported SAR was eventually organized during the 20th century which evolved to the contemporary SAR which provides services to injured, stranded, and lost people in different environments. SAR equipment, methods, and personnel can vary depending on the geography and available resources.

Unmanned Aerial Vehicle

An unmanned aerial vehicle (UAV) is a type of aircraft that does not carry human operators and crew. UAVs are also sometimes referred to as drones. In recent years, there has been a growing interest in research regarding UAV (Nex et al., 2022). Among the trends is discovering its applications to various commercial and civil tasks, and towards the development of autonomous UAVs (Alotaibi et al., 2019). For this study, the focus will be on the capability of UAVs in the field of search and rescue.

Usage of UAVs for search and rescue operations

Search and rescue (SAR) is the activity in which the goal is to aid people that are in danger or distress. This includes retrieving survivors from natural disasters in the minimum time possible (Alotaibi et al., 2019), searching for missing persons (Waharte and Trigoni, 2010), etc. In this field, time is very crucial as delays could potentially endanger lives and even lead to losses. Chiu et al. (2020) mentioned about the principle of the “golden 72h” in SAR where the first 72 hours is said to be the most critical time for victims to be rescued. Figure 1 shows how quickly the survival rate of victims deteriorate after an earthquake. One of the strategies used for a more effective SAR operation is employing the use of UAVs.

UAVs are agile, fast, and as it operates in the air, it faces less obstacles that could

Chart

Description automatically generated

Fig. 1. Survival ratio of victims during an earthquake. (Lifted from Chiu et al., 2020)

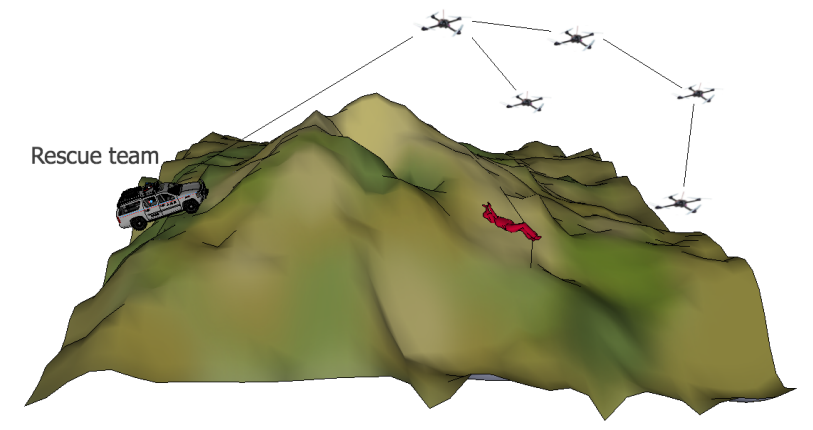


Fig. 2. An example SAR scenario employing the use of UAV to augment the searching capabilities of the rescue team. (Lifted from Waharte and Trigoni, 2010)

potentially block its path and vision. As shown in Figure 2, areas of interest can be surveyed faster and this makes the SAR operation more efficient. UAVs can provide aerial images that allow for the identification of people that may need assistance. This strategy already proven its effectiveness such as in cases where UAVs found and rescued a trapped 60-year-old hiker in Snowy Canyon State Park, Utah on 2019, a car crash victim in the U.K. on 2018, and two reportedly missing hikers in Pike National Forest in Colorado on 2017 (Dukowitz, 2019). Manually checking for victims could be difficult and prone to human errors, therefore, there were studies that used object detection techniques for finding victims through the UAV camera.

Usage of object detection techniques for human detection in search and rescue

In recent years, the development of deep learning object detection techniques is improving at a fast pace. These techniques allow for the creation of various models that is trained to detect objects with high precision. There are past studies that use those object detection techniques for human detection in the context of SAR. In the next paragraphs, various studies conducted in this field will be discussed. Table 1 summarizes the main points of the following studies.

In a study conducted by Qingqing et al. (2020), they trained a YOLOv3 model for human detection for marine search. Because of the lack of open dataset of people in water,

Table 1. Comparison of different related works for object detection in drone images.

|  |  |  |
| --- | --- | --- |
| Title | Models Used | Remarks |
| Deep Learning Approach in Aerial Imagery for Supporting Land Search and Rescue Missions (Božić-Štulić et al., 2019) | Faster R-CNN, Saliency-based RPN | 88.9% recall with 34.8% precision. Provided the HERIDAL dataset. For Faster R-CNN training, dataset was cropped into 1333 x 1000. |
| Towards Active Vision with UAVs in Marine Search and Rescue: Analyzing Human Detection at Variable Altitudes (Qingqing et al., 2020) | YOLOv3 | 98.8% correct detections at <60 meters, 83.3% correct detections above 90 meters. Inference time was not documented. IoU threshold was 0.1. |
| Real time object detection for aerial search and rescue missions for missing persons (Domozi et al., 2020) | SSD | 65% recall, 95% precision. Used HERIDAL dataset. Mentioned the importance of recall in SAR context. Dataset was cut in quarters. |
| Drone-surveillance for search and rescue in natural disaster (Mishra et al., 2020) | R-CNN, R-FCN, SSD based proposed model | 98% mAP@0.5 on their own dataset which is based on close drone shots. Recall metric is not documented. |
| Toward AI-Assisted UAV for Human Detection in Search and Rescue Missions (Rizk et al., 2021) | YOLOv3 | 78.78% mAP@0.5 on their own dataset which is based on close drone shots. Recall metric is not documented. |
| Detecting Humans in Search and Rescue Operations Based on Ensemble Learning (Dousai and Lončarić, 2022) | EfficientDET + Bi-FPN and FC-FPN | 95.11% best mAP in HERIDAL dataset. Recall metric and inference time was not documented. Cropped the dataset into 512 x 512, 640 x 640, and 1024 x 1024. Accuracy is better on lower resolution. |

they collected 458 photos with a resolution of 4000 x 2250. The images were taken in various height from 20 meters to 120 meters. They achieved mAP@0.1=69.84% on their dataset. They used 0.1 IoU as only the approximate location of persons is needed and not the exact size. In addition, they observed that at less than 60 meters, they got 98.8% correct detections while 83.3% correct detections above 90 meters. This shows that as the height of the drone increases, the objects are less likely to be detected. Another study by Rizk et al. (2021) also used YOLOv3 on a different dataset using internet images. In their testing, they got an mAP@0.5=78.78% which is relatively high, but this is given that they tested it on a captured video at a mere height of 14 meters.

In a different study conducted by Mishra et al. (2020), they trained object detection models for human detection that will be used for drone surveillance in events such as natural disaster. Specifically, they trained Faster R-CNN, R-FCN, and they also proposed their own model. The models were trained in a dataset consisting of drone shots from 10

Graphical user interface, website

Description automatically generated

Fig. 3. Images in the dataset used in the study of Mishra et al. (Lifted from Mishra et al., 2020).

meters to 40 meters with humans within the frame having various poses such as running, waving, etc. such as what is seen in Figure 3. They achieve the highest mAP@0.5=98.8% with Faster R-CNN. However, the high mAP is due to the fact that their dataset is based on close drone shots, and it can be seen that the view range of the image can be further improved.

Božić-Štulić et al. (2019) conducted a similar study to earlier mentioned literatures but in their study, they utilized a dataset called HERIDAL that has a basis on SAR operations. They also made this dataset open to everyone. In this dataset, they provided drone shots taken from a very high altitude so that it covers a wide area, which is preferred in SAR. However, the resolution of the images was also high (4000 x 3000) and objects to detect are relatively very small. This is problematic as a very high-resolution image can be very computationally expensive to train. On top of that, simply resizing the image in this case will heavily affect the detectability of the already very small objects. To address this issue, most studies that utilized this dataset cropped the images before it is fed to their object detection model. In their own study, they proposed two models: a two-step model that uses a saliency-based region proposal network and a VGG16 image classification network, and Faster R-CNN. Their proposed model achieved 88.9% recall and 34.8% precision. On the other hand, Faster R-CNN achieved 85.0% recall and 58.1% precision when the images were cropped into 1333 x 1000.

Another study utilizing the HERIDAL dataset was conducted by Domozi et al. (2020) which focuses on real time human detection for aerial SAR for missing people. They used SSD object detection model in their study, and they managed to get a recall of 65.4% and a precision of 96.4% with a speed of 11 to 17 fps in real time. In their case, they cropped the images into quarters (2000 x 1500). Moreover, they mentioned that in the context of finding humans to rescue, the recall metric is more important than precision. The other mentioned literatures failed to document such important metric in their study.

Lastly, Dousai and Lončarić (2022) used EfficientDET in their study and they used multiple resolution crops in their study. Specifically, they derived the HERIDAL dataset into 512 x 512, 640 x 640, and 1024 x 1024 resolution. They managed to achieve mAP=95.11% in their study which was achieved on the 512 x 512 resolution. This signifies that detection will be better if the detected object is relatively bigger compared to the entire image. However, they did not document the recall metric.

The Dataset

There are many available aerial datasets that are open sourced. Some of these are Vision Meets Drones (VisDrone) (Zhu et al., 2018), The UAV123 (Mueller et al., 2016), Okutama-action (Barekatain et al., 2017), HERIDAL (Božić-Štulić et al., 2019), Lacmus Drone Dataset (LADD), etc. While most of those are well labeled, this study aims for a model that is usable for SAR operations. Therefore, a dataset oriented for human detection in the SAR context is preferred. From the mentioned datasets, only HERIDAL and LADD datasets fit this criterion. For this study, the HERIDAL dataset will be used since there are past studies that utilized this dataset. With metrics from the models trained in past studies, it will allow this study to be compared to those studies to some extent, and possibly pinpoint the improvements introduced by this study.

The HERIDAL dataset was put together by Marušić et al. (2018) in their study entitled “Region Proposal Approach for Human Detection on Aerial Imagery” and it is composed of over 1500 well-labeled images for training and 101 for testing. Most images



Fig. 4. Some images in HERIDAL dataset that shows the diverse environments the human can be found. From top-left to bottom-right: the object can have a shadow, have different scales and poses, in an area camouflaged by rocks, camouflaged by trees, captured with motion blur, and have low light conditions. (Lifted from Dousai and Lončarić, 2022)

in the dataset have a 4000 x 3000 resolution with one subset that uses 4000 x 2250. The images were taken from various places in Croatia and BiH (Bosnia and Herzegovina). From different locations, students and volunteers act as missing persons and the images were captured from a height of 40 meters to 65 meters. Although the situation in the dataset is staged, the scenes are made as realistic as possible such as having different terrains, including mountains and wilderness, people having different poses and clothing, etc. It is important to highlight that there is no disturbing image in the dataset despite it being geared for search and rescue. There are multiple environmental conditions where the human to detect is present, which makes the dataset challenging as seen in Figure 4. Moreover, as the shot is taken from a very high altitude, objects appear very small, as seen in Figure 5. This is one of the important aspects that must be considered in training the object detection model.

A picture containing outdoor, mountain

Description automatically generated

Fig. 5. An example image from the HERIDAL dataset and its annotation.

Object Detection

The technique used to detect object instances of a specific class (e.g., buildings, humans, bag, etc.) is object detection. Object detection techniques can be applied to a wide range of applications, especially in computer vision. Over the past two decades, multiple object detection techniques emerged as the technology improves (Zou et al., 2023). Early techniques are based on features that are explicitly handcrafted. One of such detectors is Viola Jones Detector, which was proposed by P. Viola and M. Jones in 2001. The main application of the algorithm is for face detection. The algorithm works through the use of sliding windows that tries to find a human face in all possible locations in an image (Viola and Jones, 2001). Another would be Histograms of Oriented Gradients (HOG) which was proposed by N. Dalal and B. Triggs in 2005. This detector was a human detector that uses vectors to see the overall shape of the object (Dalal and Triggs, 2005). These models pale in comparison to the more advanced object detectors we have today. Current object detection approaches utilize machine learning. This started with the rediscovery of Convolutional Neural Networks (CNN) for image classification in 2012 by Krizhevsky et al. (2012). Their AlexNet is the winner of the Imagenet Large-Scale Visual Recognition Challenge (ILSVRC) 2012. With this is the discovery that a deep convolutional neural network can learn a robust and high-level features of an image. Eventually, the use of CNN is introduced to object detection in 2014. From that point onwards, deep learning-based object detection rapidly developed to create more precise and more robust models.

Performance metrics for object detection

Before discussing about the different object detection models, it is important to learn about the metrics used in the object detection field for measuring performance. For different researchers to compare their models to previous ones, there needs to be a similar dataset used, metrics, and the hardware used for inferencing. The most common dataset that is used by researchers is Microsoft COCO (Common Objects in Context). It is a very large dataset that contains 91 object categories, 2,500,000 labeled instances, and 328,000 images (Lin et al., 2014). In terms of hardware, the common graphics processing unit (GPU) that is used is Nvidia Tesla V100. For speed, they either document it as frames per second (fps) or inference time (in milliseconds). For the performance of the model in terms of how good it is at detecting things, the most used metric is the mean Average Precision (mAP).

mAP calculates the tradeoff between precision and recall. Precision is a measure of how much among the objects predicted by the model is correct. On the other hand, recall is a measure of how much among the correct objects did the model correctly predict. Average precision (AP) is the area under the precision-recall curve (Su et al., 2015). In object detection, one more criterion is important for calculating the average precision – the correctness of the bounding box (BB). This is done through Intersection over Union (IoU) which is the ratio of the intersection and the union of the predicted BB and ground truth BB. This can produce a value from 0 to 1 where a value closer to 1 means that the predicted BB is closer to the ground truth BB while a 0 means there is no overlap at all. An IoU of 0.5 is the usual threshold in which a prediction is considered as a true positive (assuming also correct classification). Finally, mAP is simply the average of all AP from all image classifications (Tan, 2022). mAP@.5 means the mAP of the model with the IoU threshold set to 0.5, and usually, just mAP means the average of all mAP from mAP@.5 to mAP@.95 with a step of 0.05 IoU. Table 2 summarizes the formula that is used in calculating metrics. While mAP is enough to analyze the effectiveness of an object detection model, in the context of SAR, the recall is more important as the priority is to detect all victims if possible and it is better to find more false positives (wrong detections) than false negatives (failing to detect a victim) (Domozi et al., 2020). However, it is still important to improve precision

Table 2. Summary of equations for object detection performance metrics.

|  |  |
| --- | --- |
| Description | Formula |
| Intersection over Union |  |
| Precision: ratio of TP and the total number of predicted positives |  |
| Recall: ratio of TP and the total number of ground truth positives |  |
| Average Precision: area under the precision-recall curve |  |
| Mean Average Precision: average of the AP of all classes |  |

TP: true positive (correct prediction), FP: false positive (wrong prediction), FN: false negative (undetected objects), Q: number of classes

of the model as much as possible to improve the efficiency of the object detector.

You only look once

You Only Look Once (YOLO) was proposed by J. Redmon et al. in 2015 and was the first one-stage CNN based object detector (Redmon et al., 2016). This means that the prediction of the bounding box and the class is done in a single CNN. One-stage object detectors are generally fast. That’s why they are viable for real-time detection. Over the years, it had been one of the most popular object detection models that is used, and a lot of YOLO versions have been released, some from an entirely different set of researchers.

The original YOLO (Redmon et al., 2016) works by splitting the image into grid cells with a dimension of s x s (7 is the default s value) and with each cell being responsible of detecting an object if the object’s center is inside the said cell. It achieved an mAP of 63.4% at 45 fps on the PASCAL VOC dataset. However, there are still some improvements to be done in this model such as the difficulty of the model in detecting close proximity objects as each grid has a maximum proposal of only 2.

YOLOv2 (Redmon and Farhadi, 2016) implemented multiple techniques to improve the YOLOv1 model. One of these was the introduction of batch normalization to the architecture of the model which made the training faster. Batch normalization alone already improved the mAP of the original YOLO by 2%. Moreover, the input image size for training was increased to 448 x 448 pixels from 224 x 224 pixels, which also resulted with an increase of 4% in the mAP. Anchor boxes were also introduced in this model. These are boxes of predefined dimensions which can fit the object of interest. The mAP of this model is 76.8% at 67 FPS with VGG-16 backbone and 78.6% at 40 FPS with GoogleNet backbone.

YOLOv3 (Redmon and Farhadi, 2018) is another improvement to the YOLO architecture. One problem that was observed in YOLOv2 was that it had difficulty in detecting small objects. This was a result of the focus of developing deeper networks at that time as these networks leads to a higher accuracy. However, as the image progress through the network, the progressive down sampling leads to losing some features. These loss of features affects the detectability of small objects. To address this issue, YOLOv3 used the concept of residual networks (He et al., 2015) to preserve the features from the shallow layers to the deeper layers. It achieved a 33.0% mAP on the COCO dataset.

YOLOv4 (Bochkovskiy et al., 2020) is the start of the series of YOLO versions that are not developed by the original author of YOLO. This version of YOLO utilizes a lot of object detection techniques. First, there were multiple backbones or feature extractors that was tested for YOLOv4 and CSPDarknet-53 was chosen after some experimentation. The CSP in the name stands for Cross-Stage-Partial-connections which is a technique that enhances the learning capability of CNNs (Wang et al., 2019). Feature aggregation networks were also implemented and tested for this model. These are techniques that involves aggregating features from the different depths within the network and in turn, improves the networks capability. Some of these techniques are FPN (Lin et al., 2017) and PANet (Liu et al., 2018). YOLOv4 utilized the latter. Aside from those, YOLOv4 also included more techniques such as bag-of-freebies, bag-of-specials, and much more optimizations. This model resulted in an mAP of 43.5% with a speed of 65 fps on the COCO dataset with the Tesla V100 GPU.

While YOLOv5 (Jocher, 2020) contributed similarly to YOLOv4, its main improvement is the concept of automatic anchor box learning. Anchor boxes was first implemented in YOLOv2. Because it was of set dimensions based on the COCO dataset, it showed that it does not adapt quickly to a different dataset. In YOLOv5, the anchor box selection process was included in the pipeline which accelerates the overall training process. The latest version of YOLOv5, which is v7.0, was able to reach 55.8% mAP on the COCO dataset with an inference time of 26.2 ms on a V100 GPU.

YOLOv6 (Li et al., 2022) utilized an anchor-free approach. This means that the model has a better generalization ability compared to pre-defining or learning anchor boxes as well as having a faster post-processing time. This made the model 51% faster than its anchor-based predecessors. Aside from that, the authors also improved the backbone and neck of the architecture. YOLOv6 was able to reach 52.5% mAP on the COCO dataset with 98 fps on a Tesla T4 GPU.

YOLOv7 (Wang et al., 2022) gained improved performance by implementing techniques to the architecture of the model based on network efficiency. Some of these advancements includes the implementation of E-ELAN (Wu et al., 2021) which is designed by taking into consideration other factors that affects the accuracy and speed of object detection models such as memory access cost and input/output channel ratio. It was able to achieve an mAP of 56.8% on the COCO dataset with 36 fps on a V100 GPU.

YOLOv8 was released by Ultralytics in January 2023 (Jocher et al., 2023). This iteration has a very fast inference time as it requires fewer parameters to achieve its performance. Similar to YOLOv6, YOLOv8 is anchor-free, meaning, it does not initially predict objects as a box, but rather its center. Similar to YOLOv6, this improves the speed of the model. YOLOv8 can be run in the command line as it offers a command line interface (CLI), or it can also be installed as a python package using PIP. On the COCO dataset, it achieved an mAP of 53.9% at 283.3 fps with Nvidia A100 with TensorRT.

Hussain (2023) was able to conduct a more thorough discussion on the history and improvements of YOLO versions in their paper “YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection”. For easier viewing, Table 3 shows the summary of the performance of YOLO versions from v5 to v8 tested in the COCO dataset. This table specifically compares the nano/tiny versions of the model as these are the ones that will be utilized in this study as these versions does not require too much computational power while still having good performance.

Slicing aided hyper inference

Slicing Aided Hyper Inference (SAHI) can be applied to any model during inference time. SAHI was proposed by F. C. Akyon in 2022. Motivated with the low accuracy of object detection models in drone and surveillance camera where objects appear to be small, the proponent proposed a generic solution based on slicing aided inference (Akyon et al., 2022). Figure 6 shows the overview of how SAHI is integrated in the inferencing phase of the model. The concept is simply to perform inference on smaller slices of the original model, as well as optionally taking the full inference of the model if

Table 3. Summary of the performance of YOLOv5 to YOLOv8 on the COCO dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| YOLO version | mAP | Latency | Parameters |
| YOLOv5n | 28.0 | 1.7 ms | 1.9 million |
| YOLOv6n | 35.9 | 1.2 ms | 4.3 million |
| YOLOv7-tiny | 37.4 | 2.4 ms | 6.2 million |
| YOLOv8n | 37.3 | 0.99 ms | 3.2 million |

Aside from YOLOv8n which utilizes A100 GPU, the rest is based on Tesla T4 GPU.

Diagram

Description automatically generated

Fig. 6. Overview of SAHI framework. (Lifted from Akyon et al., 2022)

there is a need to detect large objects. Overlaps can also be set by the user to prevent objects from being lost if divided into parts in different slices. In addition to the overlaps, the slice dimensions can be manually set by the user. Basically, it is a method for detecting smaller objects without the need of retraining the model. In their paper, they were able to improve various pre-trained object detection models up to 19.8% increase in mAP. Like YOLOv8, SAHI can be used through CLI or Python code. Moreover, the SAHI library also includes tools for image and dataset slicing.

The Gaps and Proposed Solution

So far, the usage of UAVs for human detection was discussed, as well as different object detection models and techniques. One of the common issues from their studies is the failure to document the recall metric of their models. While average precision is important to an object detection model, the recall metric becomes highly important in the context of SAR as the failure to detect a human that needs rescue could possibly lead to loss of life. The importance of detecting a victim far outweighs the possibility of some false positives. Moreover, in some of the studies, the drone images are taken from a not very high altitude. This means that the drone cannot cover a large area and if ever the drone will try to observe from a higher altitude, their model will face the problem of having difficulty in detecting humans that now appears small. Moreover, the studies that utilize the HERIDAL dataset were forced to crop the images in order to properly train the object detection models. This means that it is now inefficient to feed the original high-resolution image as input to the model. This defeats the purpose of having a high-resolution, high-altitude input image in the first place.

With the improvements displayed by SAHI on the mAP of object detection models on drone images with relatively small objects, it is a promising technique that may bring improvements to the performance of an object detection model geared towards drone-based human detection. The effect of the introduction of SAHI in the series of studies conducted for human detection with the HERIDAL dataset is worth investigating. Currently, SAHI has never been applied in the context of SAR, and by extension, with the HERIDAL dataset.

In this study, the researcher will train a YOLOv8 model for human detection in the context of SAR. The HERIDAL dataset will be used as it is the best dataset to represent images that can be seen in SAR operations. Moreover, the images in the dataset appear to be taken from a very high altitude which makes the line of sight wider. Furthermore, there are past object detection studies using this dataset which can be used in comparison. The model that will be used in this study is YOLOv8. SAHI will be utilized to improve the training dataset of the model, as well as allowing the model to infer on the whole high-resolution image without sacrificing the detectability of small objects. Multiple slice dimensions will be experimented upon. Recall, mAP, and inference time of the models will be documented in this study.

Summary of the Chapter

In this chapter, the recent trend of using UAVs in commercial and civil applications was highlighted, especially its use in SAR operations. With the rapid development of deep learning-based object detection, many researchers explored the capabilities of object detection models in detecting humans in the context of SAR such as during calamities or finding missing persons. These recent studies have some gaps in them such as failure to document the recall metric of their model which is highly important in the context of SAR. Lastly, the proposed method is thoroughly discussed such as the dataset, the model to be used, how the model will be evaluated, and techniques to address potential problems such as small object detection. The next chapter will discuss the methods that will be conducted in this study.

MATERIALS AND METHODS

This chapter includes all necessary information regarding the methodology of this study. All steps from the preparation of the needed training data, usage of frameworks for implementing the object detection model, and how it will be analyzed will be discussed. The hardware and software that is used will also be mentioned in this chapter. This part of the manuscript will allow other researchers to replicate the experiments done in this study.

Environment Preparation

To make the overall process simpler, the researcher will utilize open-source libraries that allow the user to focus on training the model and not bother about manually programming the model architecture from scratch. In this study, the researcher will reference two open-source libraries. These are Ultralytics YOLOv8, and SAHI library. Table 4 summarizes the links and installation of those mentioned utilities. Conda environments will be utilized to prevent possible conflicts within installed python libraries and for overall easier management. A high performance processing environment will be utilized in this study such as COARE HPC or Google Colab.

Ultralytics YOLOv8

Ultralytics has a very powerful library that offers a lot of features with regards to YOLOv8. Their library is very easy to use and has a very detailed documentation. Training, validating, predicting, and benchmarking can easily be done in with just a few lines of code. It can also be done through the command line as they offer a CLI. Moreover, there

Table 4. Open-source libraries used in the study.

|  |  |  |
| --- | --- | --- |
| Library | Link | Installation |
| Ultralytics YOLOv8 | <https://github.com/ultralytics/ultralytics> | pip install ultralytics |
| SAHI | <https://github.com/obss/sahi> | pip install -U sahi |

is no need to think too much about implementing the YOLOv8 architecture from scratch.

SAHI

With the SAHI library, it becomes very straightforward to implement Slicing Aided Hyper Inference, which is one of the strategies for small object detection as mentioned in the literatures. By simply installing their library on top of the other object detection libraries such as Ultralytics YOLOv8, sliced inference can be immediately applied to trained models just by using a function within the library. Moreover, dataset slicing can easily be done using the library.

Dataset Preparation

In this study, the HERIDAL dataset (Božić-Štulić et al., 2019) will be used to train the models. This dataset is very compatible with the objectives of the study because the dataset is composed of multiple high-altitude drone shots and was made with SAR in mind as mentioned in the literatures. The dataset is publicly available and can be accessed in this link: <http://ipsar.fesb.unist.hr/HERIDAL%20database.html>. The dataset consist of over 1500 images for training and 101 for testing. The dataset labels are stored in .xml files with the bounding boxes stored as the coordinates of the upper left and the lower right corners of the object (similar to Pascal VOC). The patches folder consists of cropped images from the testImages and trainImages folder where patches without a person in it is stored in the negative folder and patches with a person is stored in the positive folder. These specific folders will not be used in this study. Only the testImages and trainImages folder will be utilized as test sets and train sets for the model, respectively.

In order to utilize the built-in functionalities of the SAHI library such as dataset slicing and annotation conversion to YOLO, the annotation format of the images must first be converted from PASCAL VOC (Everingham et al., 2010) format to COCO (Lin et al., 2014) annotation format. PASCAL VOC stores the annotations as XML files with each image having their own XML file. This file contains some metadata such as filename, image size, and the coordinates of the bounding boxes with their respective labels. On the other hand, COCO annotation is formatted as a single JSON file for all images in a set. This file already contains all information about the dataset, the images, the corresponding annotations, and the labels. Aside from those, the YOLO annotation format is also needed by the YOLOv8 library in order to train the model. This annotation format is composed of multiple TXT files, one for each image, that simply contains the coordinates of the bounding boxes and its corresponding label. The researcher will create a python script that converts the PASCAL VOC annotation format of the HERIDAL dataset into COCO annotation format. Following this, with the built in COCO to YOLO converter of the SAHI library, the researcher will utilize that to have an annotation format to prepare the dataset for YOLOv8 training. During this step, the train set will be further divided into the train set and the validation set. 90% of the original train set will be used as the YOLO train set while the rest of the 10% will be used as the YOLO validation set. This is to ensure that the test set is a set that the model did not “see” at all during the training and reduce bias in the final evaluations. The reason why the researcher opted not to directly convert PASCAL VOC to YOLO directly is because the COCO format is important for other tasks such as dataset slicing and for evaluating the model later on.

As mentioned in the literatures, the 4000 x 3000 resolution of the dataset and the relative size of its objects is problematic, that’s why studies using the dataset employ a preprocessing step that involves cropping the images first. Fortunately, one feature of the SAHI library is dataset slicing. The slice dimensions that will be used are 512 x 512 and 640 x 640, following the top two best performing models in the study of Dousai and Lončarić (2022). The overlap to use will be either 0.1 or 0.2. 0.1 overlap is preferable as it will mean that there will be less images to process as there will be less redundancy. This will be decided upon visual inspection. If 0.1 overlap is enough to make sure that an object that becomes partially visible during the crop appears fully complete in at least one of the images, that overlap will be used. Else, a higher overlap will be utilized. Slices with no object will be discarded.

Training The Model

With the dataset ready, the researcher can now proceed to training the model. The model that the researcher plan to use is YOLOv8 as it showed better performance on the other YOLO models while not having too many parameters, which can make the model faster overall. Moreover, the variant that will be utilized will be the nano (i.e., YOLOv8n) as it is lightweight and does not take too much computing resources to train. Moreover, while it is not the best when it comes to detection accuracy among the variants, it is the fastest and this will be a more practical choice when it is taken into account the number of times the model will be run in a single image because of the multiple slices. This multiple inferencing in one image amplifies the delay caused by the object detection model in addition to the cropping delay. Furthermore, multiple experiments will be done for training for the sake of hyperparameter optimization in hopes of making the model achieve better performance such as higher mAP. There are multiple hyperparameter combinations that will be tested by the researcher. Among these are epochs, batch size, image size, and optimizer.

Evaluating The Model

The trained object detection models will be evaluated in their performance in detecting the objects in the test set of the dataset. In this study, different SAHI slice dimensions will be evaluated. Specifically, 320 x 320, 640 x 640, and 1280 x 1280. In all test cases, a 0.1 or 0.2 overlap will be utilized, depending on what was used during the dataset slicing phase as the researcher plans on making the overlap the same for the same reason mentioned in the dataset preparation section. Specific metrics that will be documented during the implementation include recall, precision, mAP, and the inference time of the models. Multiple metrics will be documented in order to have a metric to compare to past studies as some utilize different metrics in their discussion. These calculations are discussed in more detail in chapter 2, but the libraries mentioned already have these functionalities already. Just like what is mentioned in the past chapter, a high recall is very important for a human detection used for SAR, a high mAP means the model is generally good in detecting objects, and a low inference time is preferred when processing power is limited. These metrics will help with evaluating the usability of the models in specific situations.

Overall Framework

Figure 7 summarizes the overall process that will be conducted for this study. The environment that will be used for training and evaluating models will be created with the help of conda. Ultralytics YOLOv8, and SAHI libraries will be installed. The next step is to prepare the dataset to be used in the study, which is the HERIDAL dataset sliced into 640 x 640 patches with 0.2 overlap using a functionality in the SAHI library. After that, a YOLOv8 model will be trained on the dataset. Lastly, the models will be evaluated by testing it on the full resolution HERIDAL test set. This evaluation utilizes SAHI with the slice dimensions of 320 x 320, 640 x 640, and 1280 x 1280, all with 0.2 overlap. Recall, mAP, and inference time of each inference will be documented for comparison and discussion.

Diagram

Description automatically generated

Fig. 7. Overall methodology process.

LITERATURE CITED

Akyon, F.C., Altinuc, S.O., Temizel, A., 2022. Slicing Aided Hyper Inference and Fine-tuning for Small Object Detection, in: 2022 IEEE International Conference on Image Processing (ICIP). pp. 966–970. https://doi.org/10.1109/ICIP46576.2022.9897990

Alotaibi, E.T., Alqefari, S.S., Koubaa, A., 2019. LSAR: Multi-UAV Collaboration for Search and Rescue Missions. IEEE Access 7, 55817–55832. https://doi.org/10.1109/ACCESS.2019.2912306

Barekatain, M., Marti, M., Shih, H.-F., Murray, S., Nakayama, K., Matsuo, Y., Prendinger, H., 2017. Okutama-Action: An Aerial View Video Dataset for Concurrent Human Action Detection. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 28–35.

Bevacqua, G., Cacace, J., Finzi, A., Lippiello, V., 2015. Mixed-Initiative Planning and Execution for Multiple Drones in Search and Rescue Missions.

Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M., 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. https://doi.org/10.48550/arXiv.2004.10934

Božić-Štulić, D., Marušić, Ž., Gotovac, S., 2019. Deep Learning Approach in Aerial Imagery for Supporting Land Search and Rescue Missions. Int J Comput Vis 127, 1256–1278. https://doi.org/10.1007/s11263-019-01177-1

Chen, G., Wang, H., Chen, K., Li, Z., Song, Z., Liu, Y., Chen, W., Knoll, A., 2022. A Survey of the Four Pillars for Small Object Detection: Multiscale Representation, Contextual Information, Super-Resolution, and Region Proposal. IEEE Transactions on Systems, Man, and Cybernetics: Systems 52, 936–953. https://doi.org/10.1109/TSMC.2020.3005231

Chiu, Y.-Y., Omura, H., Chen, H.-E., Chen, S.-C., 2020. Indicators for Post-Disaster Search and Rescue Efficiency Developed Using Progressive Death Tolls. Sustainability 12, 8262. https://doi.org/10.3390/su12198262

Cooper, D.C., 2005. Fundamentals of search and rescue. Jones & Bartlett Learning.

Dalal, N., Triggs, B., 2005. Histograms of Oriented Gradients for Human Detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05). Presented at the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), IEEE, San Diego, CA, USA, pp. 886–893. https://doi.org/10.1109/CVPR.2005.177

Domozi, Z., Stojcsics, D., Benhamida, A., Kozlovszky, M., Andras, M., 2020. Real time object detection for aerial search and rescue missions for missing persons. https://doi.org/10.1109/SoSE50414.2020.9130475

Dousai, N.M.K., Lončarić, S., 2022. Detecting Humans in Search and Rescue Operations Based on Ensemble Learning. IEEE Access 10, 26481–26492. https://doi.org/10.1109/ACCESS.2022.3156903

Dukowitz, Z., 2019. Drones in Search and Rescue: 5 Stories of Drones Helping Save Lives [WWW Document]. UAV Coach. URL https://uavcoach.com/search-and-rescue-drones/ (accessed 5.27.23).

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The Pascal Visual Object Classes (VOC) Challenge. Int J Comput Vis 88, 303–338. https://doi.org/10.1007/s11263-009-0275-4

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. https://doi.org/10.48550/arXiv.1512.03385

Hussain, M., 2023. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. Machines 11, 677. https://doi.org/10.3390/machines11070677

Jocher, G., 2020. YOLOv5 by Ultralytics. https://doi.org/10.5281/zenodo.3908559

Jocher, G., Chaurasia, A., Qiu, J., 2023. YOLO by Ultralytics.

Krizhevsky, A., Sutskever, I., Hinton, G., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems 25. https://doi.org/10.1145/3065386

Li, C., Li, Lulu, Jiang, H., Weng, K., Geng, Y., Li, Liang, Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X., Wei, Xiaoming, Wei, Xiaolin, 2022. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. https://doi.org/10.48550/arXiv.2209.02976

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature Pyramid Networks for Object Detection. https://doi.org/10.48550/arXiv.1612.03144

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: Common Objects in Context, in: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 740–755. https://doi.org/10.1007/978-3-319-10602-1\_48

Liu, S., Qi, L., Qin, H., Shi, J., Jia, J., 2018. Path Aggregation Network for Instance Segmentation. https://doi.org/10.48550/arXiv.1803.01534

Liu, Y., Sun, P., Wergeles, N., Shang, Y., 2021. A survey and performance evaluation of deep learning methods for small object detection. Expert Systems with Applications 172, 114602. https://doi.org/10.1016/j.eswa.2021.114602

Marušić, Ž., Božić-Štulić, D., Gotovac, S., Marušić, T., 2018. Region Proposal Approach for Human Detection on Aerial Imagery, in: 2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech). Presented at the 2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech), pp. 1–6.

Mishra, B., Garg, D., Narang, P., Mishra, V., 2020. Drone-surveillance for search and rescue in natural disaster. Computer Communications 156, 1–10. https://doi.org/10.1016/j.comcom.2020.03.012

Mueller, M., Smith, N., Ghanem, B., 2016. A Benchmark and Simulator for UAV Tracking, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 445–461. https://doi.org/10.1007/978-3-319-46448-0\_27

Nex, F., Armenakis, C., Cramer, M., Cucci, D.A., Gerke, M., Honkavaara, E., Kukko, A., Persello, C., Skaloud, J., 2022. UAV in the advent of the twenties: Where we stand and what is next. ISPRS Journal of Photogrammetry and Remote Sensing 184, 215–242. https://doi.org/10.1016/j.isprsjprs.2021.12.006

Qingqing, L., Taipalmaa, J., Queralta, J.P., Gia, T.N., Gabbouj, M., Tenhunen, H., Raitoharju, J., Westerlund, T., 2020. Towards Active Vision with UAVs in Marine Search and Rescue: Analyzing Human Detection at Variable Altitudes, in: 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR). Presented at the 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), IEEE, Abu Dhabi, United Arab Emirates, pp. 65–70. https://doi.org/10.1109/SSRR50563.2020.9292596

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection.

Redmon, J., Farhadi, A., 2018. YOLOv3: An Incremental Improvement. https://doi.org/10.48550/arXiv.1804.02767

Redmon, J., Farhadi, A., 2016. YOLO9000: Better, Faster, Stronger. https://doi.org/10.48550/arXiv.1612.08242

Rizk, M., Slim, F., Charara, J., 2021. Toward AI-Assisted UAV for Human Detection in Search and Rescue Missions, in: 2021 International Conference on Decision Aid Sciences and Application (DASA). Presented at the 2021 International Conference on Decision Aid Sciences and Application (DASA), pp. 781–786. https://doi.org/10.1109/DASA53625.2021.9682412

Scherer, J., Yahyanejad, S., Hayat, S., Yanmaz, E., Andre, T., Khan, A., Vukadinovic, V., Bettstetter, C., Hellwagner, H., Rinner, B., 2015. An Autonomous Multi-UAV System for Search and Rescue, in: Proceedings of the First Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use. Presented at the MobiSys’15: The 13th Annual International Conference on Mobile Systems, Applications, and Services, ACM, Florence Italy, pp. 33–38. https://doi.org/10.1145/2750675.2750683

Silvagni, M., Tonoli, A., Zenerino, E., Chiaberge, M., 2017. Multipurpose UAV for search and rescue operations in mountain avalanche events. Geomatics, Natural Hazards and Risk 8, 18–33. https://doi.org/10.1080/19475705.2016.1238852

Su, W., Yuan, Y., Zhu, M., 2015. A Relationship between the Average Precision and the Area Under the ROC Curve, in: Proceedings of the 2015 International Conference on The Theory of Information Retrieval. Presented at the ICTIR ’15: ACM SIGIR International Conference on the Theory of Information Retrieval, ACM, Northampton Massachusetts USA, pp. 349–352. https://doi.org/10.1145/2808194.2809481

Tan, R.J., 2022. Breaking down Mean Average Precision (mAP) [WWW Document]. Medium. URL https://towardsdatascience.com/breaking-down-mean-average-precision-map-ae462f623a52 (accessed 5.27.23).

Tripicchio, P., Satler, M., Dabisias, G., Ruffaldi, E., Avizzano, C.A., 2015. Towards Smart Farming and Sustainable Agriculture with Drones, in: 2015 International Conference on Intelligent Environments. Presented at the 2015 International Conference on Intelligent Environments, pp. 140–143. https://doi.org/10.1109/IE.2015.29

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. Presented at the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, IEEE Comput. Soc, Kauai, HI, USA, p. I-511-I–518. https://doi.org/10.1109/CVPR.2001.990517

Waharte, S., Trigoni, N., 2010. Supporting Search and Rescue Operations with UAVs, in: 2010 International Conference on Emerging Security Technologies. Presented at the 2010 International Conference on Emerging Security Technologies (EST), IEEE, Canterbury, TBD, United Kingdom, pp. 142–147. https://doi.org/10.1109/EST.2010.31

Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y.M., 2022. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors.

Wang, C.-Y., Liao, H.-Y.M., Yeh, I.-H., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., 2019. CSPNet: A New Backbone that can Enhance Learning Capability of CNN. https://doi.org/10.48550/arXiv.1911.11929

Wu, W., Zhao, Y., Xu, Y., Tan, X., He, D., Zou, Z., Ye, J., Li, Y., Yao, M., Dong, Z., Shi, Y., 2021. DSANet: Dynamic Segment Aggregation Network for Video-Level Representation Learning.

Zhu, P., Wen, L., Bian, X., Ling, H., Hu, Q., 2018. Vision Meets Drones: A Challenge. https://doi.org/10.48550/arXiv.1804.07437

Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J., 2023. Object Detection in 20 Years: A Survey. Proceedings of the IEEE 111, 257–276. https://doi.org/10.1109/JPROC.2023.3238524

SCHEDULE OF ACTIVITIES

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Activities | May | | | | Jun | | | | Jul | | | | Aug | | | | Sept | | | | Oct | | | | Nov | | | | Dec | | | | Jan | | | | Feb | | | | Mar | | | |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Proposal Writing |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Proposal Defense |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Execution of Methodology |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Manuscript Writing |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Final Defense |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Activities | Apr | | | | May | | | | Jun | | | |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Proposal Writing |  |  |  |  |  |  |  |  |  |  |  |  |
| Proposal Defense |  |  |  |  |  |  |  |  |  |  |  |  |
| Execution of Methodology |  |  |  |  |  |  |  |  |  |  |  |  |
| Manuscript Writing |  |  |  |  |  |  |  |  |  |  |  |  |
| Final Defense |  |  |  |  |  |  |  |  |  |  |  |  |