Pre-Shot Sniper Detection

Jan Meijers
Studentnummer : 836897315

Arne Lescrauwaet
Studentnummer : 852617312

Abstract

Enemy snipers are a growing threat to army forces worldwide. United States Department of Defence (DOD) strives to equip soldiers with effective and overall anti-sniper detection systems. The majority of these systems however are merely based on radar, infrared (thermal), and sound signaling processing to detect and locate snipers after a sniper rifle is fired. In this study, DOD aims to develop a YOLOv8-based prototype Pre-Shot sniper detection system based on (streaming) images, to extend their knowledge of how to effectively detect and locate snipers before a sniper rifle is fired.

Our approach involved comparing two model versions. Each model underwent progressive training with increasing training set size until a DOD-required accuracy threshold was reached. Comparative analysis between the two model variants revealed no significant differences in performance. Our investigation into background environments showed that the model performed the best in a rural environment and was less effective in a snowy environment. In terms of bounding box dimensions, we observed a direct positive correlation with mAP50, though larger sizes adversely affected the mAP50-90 metric. The study concluded with successful streaming video inferences, showcasing the prototype Pre-Shot model's effectiveness.

The prototype Pre-Shot Sniper detection system demonstrated proficiency in sniper detection across diverse scenarios, achieving a mean average precision (mAP) exceeding 85% in both military and non-military environments. This includes effectiveness in (un)camouflaged settings, spanning rural, desert, snowy and urban landscapes. This work has important implications for addressing anti-sniper detection capabilities. We encourage DOD to test this prototype in real-world scenarios.

Introduction

Background

Enemy snipers are a growing threat to army forces worldwide. Snipers, armed only with a rifle, have rapidly become one of the deadliest threats (NBC news, 2006). As the war in Ukraine continues, sniper teams hunting down and killing Russian soldiers. These so-called "sniper ghosts" do have a psychological effect on their enemy (BBC news, 2023).

A sniper is a military/paramilitary marksman who engages targets from positions of concealment or at distances exceeding the target's detection capabilities. Snipers generally have specialized training and are equipped with high-precision rifles and high-magnification optics, and often also serve as scouts/observers feeding tactical information back to their units or command headquarters (Sniper, 2024).

Snipers play an important role, especially in (urban) warfare, where the challenges for snipers and counter-snipers factored heavily in battles. The use of snipers has changed with technological development since the end of World War II. Sharp shooting in contemporary technology-intensive wars requires pro-activeness and has also commenced the development of systems for finding the counter-snipers positioned against snipers (Barişik & Baltacıoġlu, 2014).

Several initiatives by the US Department of Defence (DOD) have led to the design of many (also commercially) available anti-sniper systems, merely based on radar, infrared (thermal), and sound signaling processing. The majority of these systems however detect snipers after a sniper weapon is fired. Currently, no anti-sniper system exists that provides a seamless response to a so-called Pre-Shot sniper i.e. detection of a sniper before a sniper weapon is fired.

In accordance with (US Ministry of Defense Patent, 2015), DOD strives to equip soldiers with effective and overall anti-sniper detection capabilities using state-of-the-art technologies. For this purpose and application, DOD decided to study the feasibility of the latest version of YOLO (You Only Look Once) algorithm, which is in general a single neural network that predicts bounding boxes and class probabilities directly from full (streaming) images in one evaluation (Redmon, Divvala, Girshick, & Farhadi, 2016).

Related Work

Because of security reasons little is published by DOD concerning anti-sniper systems. A research field that is most closely related to the anti-sniper detection domain - using YOLO computer vision technology - is weapon detection in the civil surveillance domain.

Weapon detection in civil surveillance, using Artificial Intelligence (AI), has gained significant attention in recent years, and numerous research studies have been conducted in this area. Computer vision technology has been used for real-time weapon detection, and YOLO object detection algorithm has emerged as a popular and efficient technique for this purpose, but further research and development are needed to improve the accuracy and reliability of weapon detection systems to ensure their responsible deployment in real-world settings (Murugan et al., 2023).

A crucial part of any application is to have desired and suitable datasets in order to train the machine learning models (Narejo, Pandey, Esenarro Vargas, Rodriguez, & Anjum, 2021). (Yadav, Gupta, & Sharma, 2023) evaluated 8 different publicly available weapon detection datasets originated from 2016 to 2021 that can be used to classify weapons. They conclude there are still certain challenges in the field of weapons detection that need to be addressed, such as a lack of (large and well-balanced) datasets, the detection of weapons in a variety of lighting conditions as well as that only a few researchers have tackled the subject of partial occlusion of weapon. Another significant problem is the capability of detecting different kinds of weapons and the detection of small-sized weapons.

Murugan et al. (2023) remarks in addition that there is no standard dataset for weapons detection and recognition. They emphasize the importance of quantity and quality of the training data to achieve high accuracy. They trained a YOLO model using a dataset of 17k images annotated weapon images with bounding boxes containing more than 8 classes i.e. firearms, knives, and other weapons. They achieved higher accuracy in removing irrelevant objects in each weapon image. Also, to use these techniques for weapon detection, it is necessary to obtain a dataset of images that includes a variety of different types of weapons in various contexts and in different lighting conditions and backgrounds according to the content. In addition, they stated that the dataset should include images of non-weapons to help the algorithm learn to distinguish weapons and other objects.

Performance metrics are key tools to evaluate the accuracy and efficiency of Pre-Shot sniper detection models. Some metrics that are not only important to YOLOv8 but are broadly applicable across different object detection models as noted by Ultralytics Performance Metrics (2023) i.e.:

1. Intersection over Union (IoU): IoU is a measure that quantifies the overlap between a predicted bounding box and a ground truth bounding box. It plays a fundamental role in evaluating the accuracy of object localization.
2. Average Precision (AP): AP computes the area under the precision-recall curve, providing a single value that encapsulates the model's precision and recall performance.
3. Mean Average Precision (mAP): mAP extends the concept of AP by calculating the average AP values across multiple object classes. This is useful in multi-class object detection scenarios to provide a comprehensive evaluation of the model's performance.
4. mAP50-95: The average of the mean average precision calculated at varying IoU thresholds, ranging from 0.50 to 0.95. It gives a comprehensive view of the model's performance across different levels of detection difficulty.
5. Precision and Recall: Precision quantifies the proportion of true positives among all positive predictions, assessing the model's capability to avoid false positives. On the other hand, Recall calculates the proportion of true positives among all actual positives, measuring the model's ability to detect all instances of a class.
6. F1 Score: The F1 Score is the harmonic mean of precision and recall, providing a balanced assessment of a model's performance while considering both false positives and false negatives.

The most common metric used for object detection models like YOLO is the mean average precision (mAP). This metric is suitable for a broad assessment of model performance. This metric considers the precision and recall of the model's predictions and provides a single numerical score for the model's performance (Terven & Cordova-Esparza, 2023). In addition to evaluating the performance of the model on the test set, it is important to analyze the types of errors the model makes to understand where improvements can be made. This can be done by analyzing the confusion matrix, which shows the number of true positives, false positives, true negatives, and false negatives for each class (Bhatti, Khan, Aslam, & Fiaz, 2021).

Knowlegde Gaps

The related work section reveals that many studies on detecting weapons use public (labeled) datasets with images of all kinds of fire weapons (i.e. guns, pistols, rifles), which are -in general- not very suitable for the Pre-Shot detection problem. Mainly because they do not represent typical and substantial sniper rifles and - most importantly - they do not represent parts (partial occlusion) of a sniper rifle. This, however, is a very important aspect of an anti-sniper detection system, because sniper and/or sniper rifles are mostly not visible as a whole.

The main problem with snipers is the gap or challenge to effectively detect real-time enemy snipers before they (can) fire their deadly weapons and kill soldiers and/or civilians instantly. This is especially true in rural environments but also in urban environments.

Snipers' most important equipment is the sniper weapon i.e. sniper rifle. The most obvious Pre-Shot detectable target(s) snipers are most probably in detecting uncamouflaged (parts of a) sniper equipment i.e. the weapon itself, in a particular part of a sniper rifle i.e. the rifle barrel of the weapon, any rifle barrel muzzle break, rifle suppressors/silencers and/or in most cases any (part of a) riflescope. Sniper rifles almost naturally come installed with attributes like barrel, muzzle break, suppressors, and/or riflescope. A sniper rifle without a riflescope is not less a sniper rifle, but less effective. Also, detecting a sniper rifle in military or hostile circumstances almost inherently detects or expects a sniper presence and/or sniper location as well.

Detecting Pre-Shot snipers based on the presence of sniper rifles is challenging. There is a lack of useful and purposefully labeled images, which can contribute to the detection of Pre-Shot snipers using computer vision methods. Although many public image datasets do exist, none of them consist of specific and/or applicable labeled images in detecting (parts of) sniper rifles such as barrels, muzzle breaks, suppressors, and/or riflescopes.

Design Research Goal

In this study, the DOD aims to construct a prototype Pre-Shot sniper detection and a high-quality custom sniper dataset, to extend their knowledge of how to effectively detect snipers in rural, snow, desert, and urban environments, before a sniper rifle is fired using state-of-the-art detection algorithms, with a minimum goal of $> 85\%$ mean average accuracy (mAP) in testing. The results of this research (including performance measurements) are reported in a so-called DOD protocol for future anti-sniper system studies and developments. For this purpose, we formulate the following main design research question:

How to construct a YOLOv8 based Pre-Shot sniper detection prototype and a high-quality custom sniper image dataset that satisfies a fast and accurate detection (i.e. classification and localization) of snipers before a sniper rifle is fired, so that army forces can save more lives of soldiers and/or civilians in the growing threat of enemy snipers in different rural, snow, dessert and urban backgrounds and/or battlefield conditions?

Design Research Objectives

To investigate the main design research question the following sub-research questions are relevant:

1. What is the minimum number of images required for training the Pre-Shot anti-sniper detection system to a predetermined threshold?

2. What is the influence of the variability of the background? Is the sniper easier to notice in certain backgrounds?

3. Does it matter how tight the bounding box is? Should one spend the extra time to get the bounding boxes just right, or is it better to annotate more images, but care less about the bounding box?

Methods

Data Collection and Pre-Processing

Based on the related work section, we concluded that custom sniper datasets for research purposes are not publicly available. Because of this, we had to build our custom sniper dataset. To fulfill our research goals, the most important step in building a Pre-Shot sniper detection system is collecting enough Pre-Shot high-quality images because it is the basic stone of building that system. We experienced that free high-quality sniper images are sparse. Although not used, massive high-quality sniper images are behind paywalls, which in general could contribute to the variability and higher resolution images.

Pre-Shot sniper images were collected from public sources on the Internet using Python-based scraping tools and the Chrome Image Downloader, based on different keywords related to snipers and/or sniper rifles (i.e. sniper snow, sniper camouflaged, etc.). For this purpose, we collected more than 1000 sniper(rifle) images. From these, we selected images by hand, based on the following criteria:

1. Sniper images with variability in the desert, snow, rural, and urban environmental backgrounds;

2. In addition to 1, sniper images of completely and partly visible or occluded sniper rifles in different colors, types, sizes, angles, profiles, and light conditions;

3. In addition to 2, sniper images of sniper rifle parts i.e. rifle scopes, barrels, suppressors, and muzzle breaks of different sizes, colors, front/back and sideways angles, and partly visible or occluded;

4. In addition to 3, sniper images with (partly) camouflaged sniper men/women and/or camouflaged sniper rifles or camouflaged parts of sniper rifles;

5. In addition to 4, sniper images of snipers in different positions (standing, laying kneeling, partly visible or hidden, with and without camouflage suites);

6. In addition to 5, sniper images closeup images, and far away/distance positions.

As a result of the application of these stringent criteria, we managed to deliver a reasonably high quality, balanced, and high variability Pre-Shot sniper image dataset collection, consisting of 711 images. Figure 1 shows some examples of selected images.

Figure 1. Examples Pre-Shot Sniper Images

Building Labelled Pre-Shot Sniper Dataset

For labeling and annotation of Pre-Shot sniper images we used Roboflow because this web-based tool is free to use without any limitations/restrictions for uploading a massive number of images and allows annotation processing. The selected 711 Pre-Shot sniper images were uploaded to Roboflow to label and annotate the images. Roboflow divided the labeled images automatically and randomly into 3 parts: training set (70%), validation set (20%), and test set (10%). All images were annotated by bounding boxes to one "snipers" class, as shown by some examples in Figure 2.

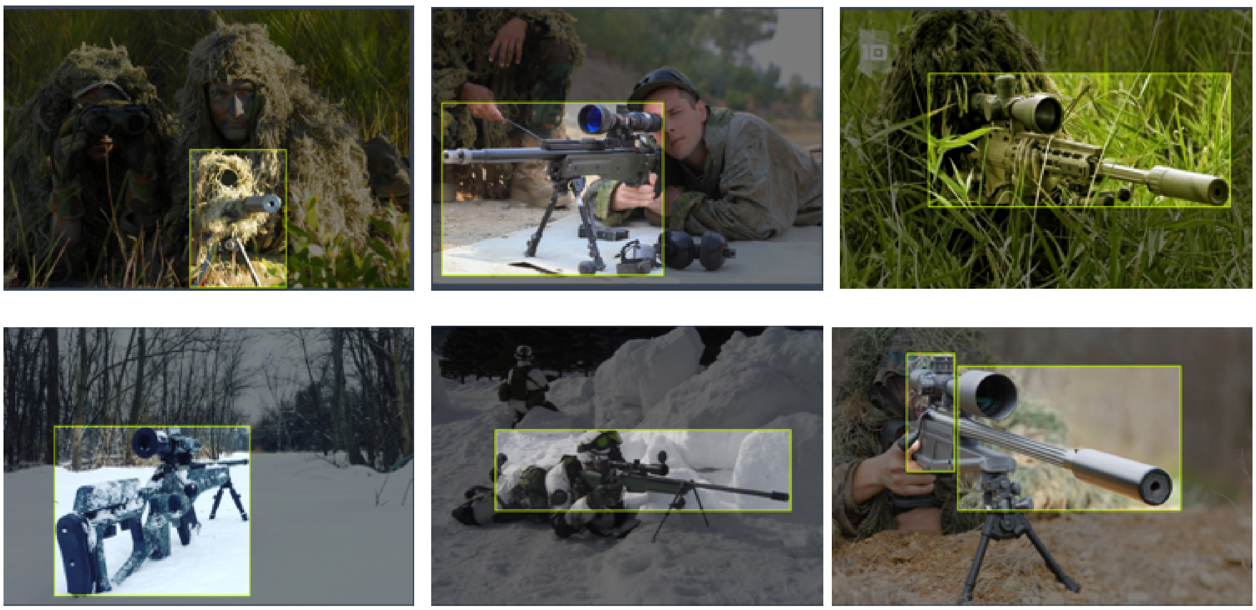


Figure 2. Examples of Annotated Pre-Shot Sniper Images

Training the Pre-Shot Sniper Detector

The pre-Shot sniper detection algorithm is based on the pre-trained (on COCO) YOLOv8 model made by Ultralytics. The model is trained for 100 epochs with a confidence of 0.25 and an image size of 640. For training, validating, and testing purposes, we used a local GPU Nvidia RTX3080 (10GB) 16 GB RAM CPU Intel i5-13600 KF based computer.

To answer the first sub-research question concerning the minimum number of images, the model was first fine-tuned on 100 images, this amount was iteratively incremented by 50 images until we could reach the DOD goal of $> 85\%$ mean average accuracy (mAP) in testing.

In addressing the sub-research question regarding the impact of bounding box size, we examined its relationship with mean average precision (mAP). Our analysis focused on evaluating how bounding box dimensions influence both mAP and mAP50-95 metrics. This was assessed by implementing bounding box augmentation strategies, as detailed in Figure 3. For augmentation, we employed a factor of 0.2, effectively enlarging the bounding box dimensions (both width and height) significantly, yet without excessively dominating the image space.

In exploring the sub-research question about the effect of background variability, our investigation focused on identifying potential differences in distribution and bias across three distinct

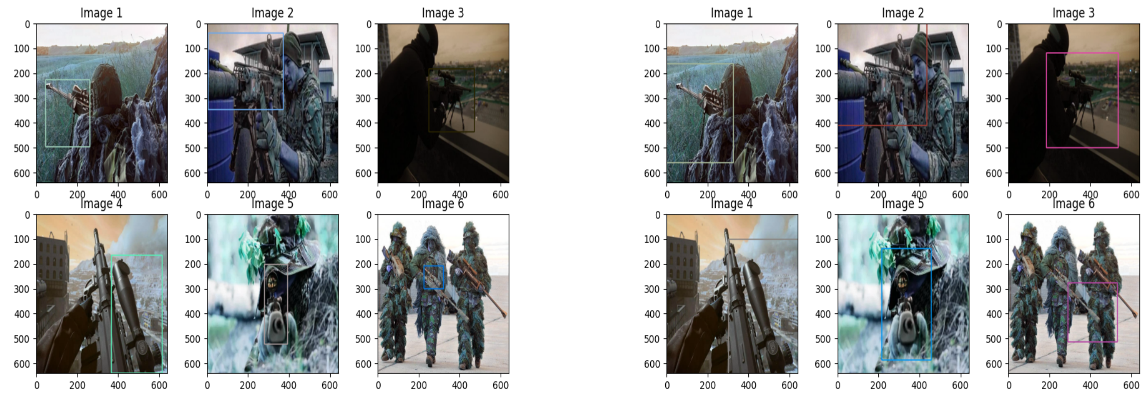


Figure 3. Original vs Augmented Bounding Boxes

environmental backgrounds: snow, desert, and rural. To analyze the normality of distribution in these categories, we employed the Shapiro-Wilk test (Shapiro-Wilk test, 2024), supplemented by Barlett's test (Bartlett's test, 2024) for assessing variance equality. Our samples comprised 30 images from each category. These samples were estimated (power analysis) with respect to our knowledge of building the custom Pre-Shot Sniper dataset population. To determine any significant differences in the metrics observed across these backgrounds, we applied Welch's ANOVA test (ANOVA test, 2024).

Inference on YouTube videos

This inference experiment was used to simulate the Pre-Shot Sniper detector in a more realistic setting. In a real-world scenario, the Pre-Shot Sniper detector will constantly be fed at high speed with a special and very high quality/resolution camera-sensor system. This camera system will be part of the military or soldier equipment but was not in scope in this study. We selected different and more or less realistic short YouTube sniper movies to test the detection performance of the Pre-Shot sniper detector. We emphasize that the Pre-Shot Sniper detector should be evaluated in real-world scenarios, varying enemy sniper distances, varying environments, and different weather conditions (i.e. snowing, raining, stormy/windy, misty, and twilight) with the application of the aforementioned DOD camera.

Results

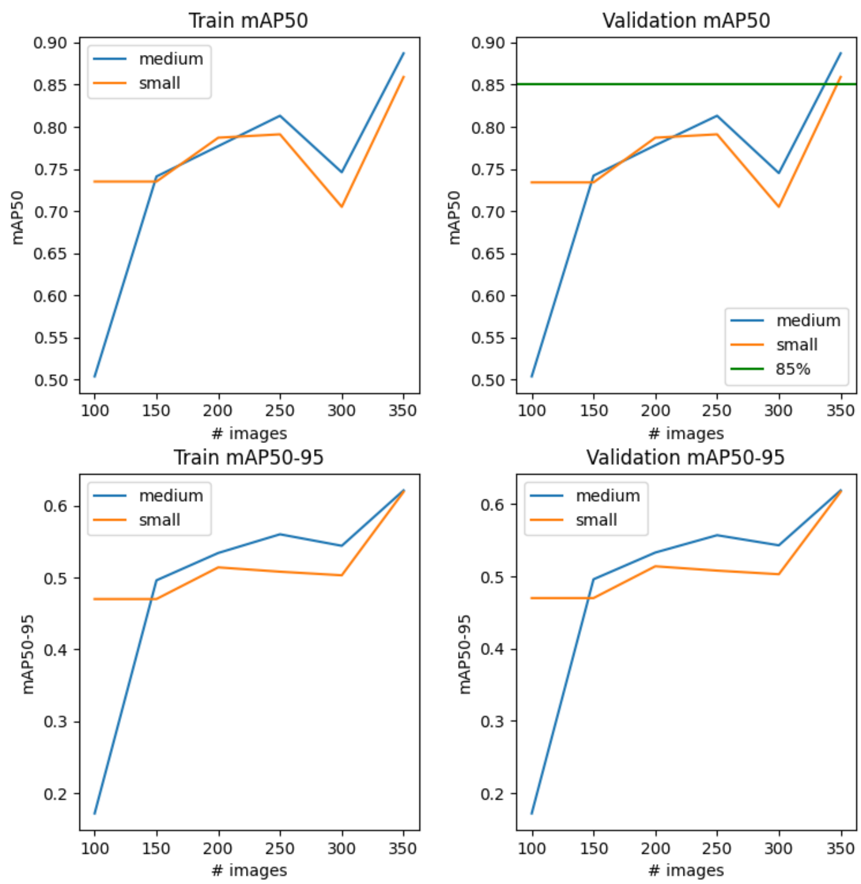
Minimum number of images

Figure 4 showcases the results of the models fine-tuned using the small and medium weights, specifically the mean average precision (mAP) and mean average precision within the 50-95% Intersection over Union thresholds (mAP50-95) in both training and validation on various quantities of images.

A comparative analysis between the small and medium weights variants reveals striking similarities. Both variants achieve similar performance, the small weights model seems to do better with fewer images than the medium weights. However, the medium weights model has a slight advantage over the small weights variant when trained on more images achieving a slightly higher mAP50 in both training and validation.

The study's findings conclude that 350 images are necessary to attain the desired 85% mAP50 threshold in validation. However, the mAP50-95, standing at 62%, indicates a notable discrepancy in the accuracy of bounding box placements. While the detection system successfully identifies the presence of snipers with bounding boxes delineating the general location, achieving precise localization remains a challenge. This discrepancy between detection and localization accuracy is likely attributed to the annotation methodology, which was a combination of the rifle and the sniper or just the rifle. Had we focused solely on annotating the sniper rifle itself or only on the combination of the sniper and his rifle, which was often not possible, there is a possibility that we could have achieved a higher mAP50-95 score.

Figure 4. mAP50 and mAP50-95 in training and validation

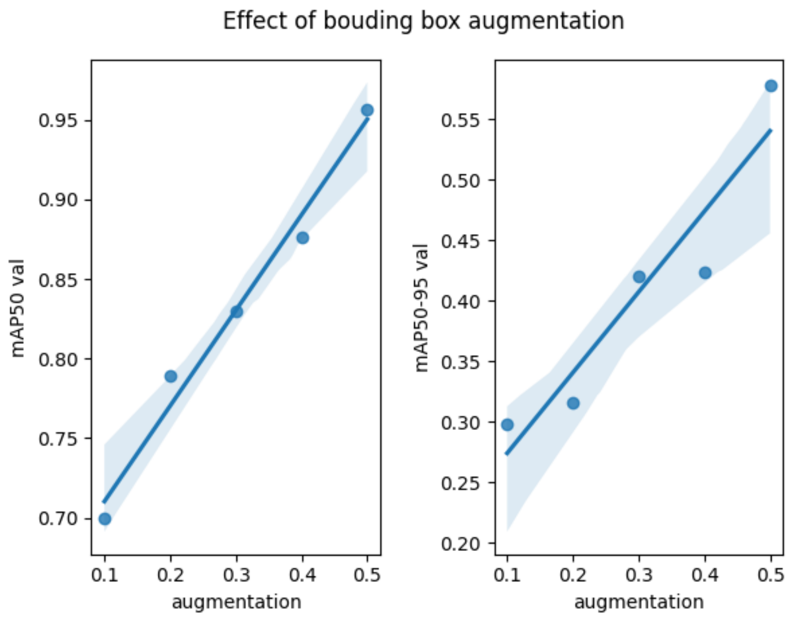


Influence of the bounding box

The results from our experiments demonstrate a clear correlation between bounding box size and mean average precision (mAP). There was a distinct increase in mAP observed with the expansion of bounding box dimensions. This increase was more pronounced when contrasted with scenarios that did not involve bounding box augmentation, particularly in small and medium-weight configurations.

However, it's essential to note that a significant reduction in mAP50-95 offsets this improvement in mAP50. Although larger bounding boxes yield noticeable benefits, this approach could be challenging in situations involving closely aligned objects or when high precision in bounding box localization is required. These findings emphasize the delicate balance involved in bounding box augmentation strategies and the importance of adapting these strategies to the specific requirements of the object detection task. A detailed comparison of the metrics for small and medium weights is presented in Figure 4.

Figure 5. influence of bounding box size on mAP and mAP50-95



Influence of the background

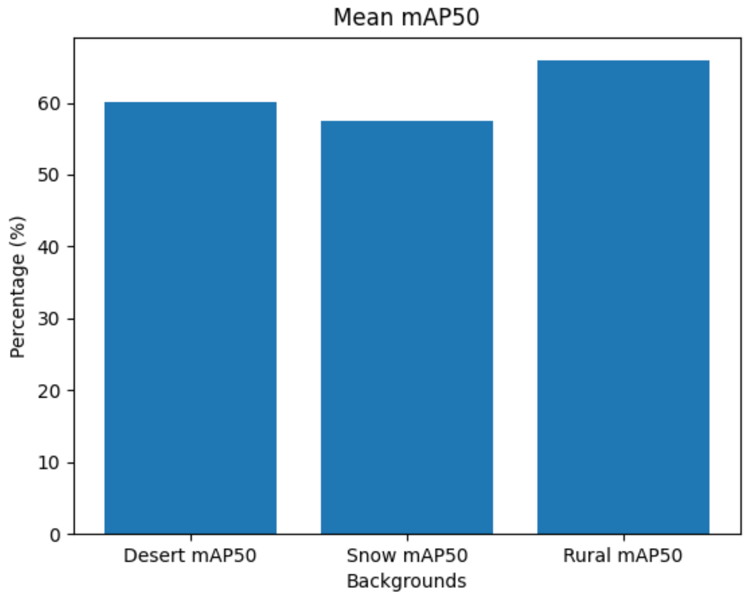
In assessing potential distinctions among our three background categories: "snow, desert, and rural", a Welch's ANOVA test was employed to examine for significant differences in the validation metrics of predictions on those backgrounds. To apply the ANOVA test appropriately, certain assumptions regarding the data needed to be satisfied, namely: the data are independent, normally distributed, and homogeneity of variance among samples.

The entire sample generation procedure can be found in the "compare_bias.ipynb" file in the "generate_sample()" function. Regarding the assumption of normality, the Shapiro-Wilk test was employed. The values for the different backgrounds were close to 1 indicating that they all had a normal distribution.

To assess the similarity of variances among different backgrounds, Bartlett's test was employed. The Barlett statistic was quite high (9.38) and the p-value was smaller than 5% thus the hypothesis that the population variances of the three groups are equal is rejected. This means that we need to use a Welch's ANOVA test.

The outcome of Welch's ANOVA test indicates a significant difference in the background pairings. Looking at the mean accuracy of the mAP50 scores, see Figure 6, We can see that the model scored best on the rural background and worst on the snow background.

Figure 6. mean mAP50 of background samples



Inference on videos

To assess our trained model in a more realistic setting, we tested it on various videos (1, 2,3) using the default prediction settings. Our findings suggest that the model performs reasonably well in spotting snipers, even those partially hidden (see Figure 7). Snipers that are well-camouflaged or very far away can still be rather difficult to detect resulting in false negatives.

Figure 7. Detection of hidden snipers

(a) Hidden sniper 1

(b) Hidden sniper 2

However, we've noticed some inaccuracies in the model detections, as depicted in Figure 8. The model seems to focus heavily on the barrel when detecting snipers, thus objects resembling a rifle barrel are suspect to false positives such as image "a" in Figure 8. These inaccuracies can be addressed by adjusting detection parameters, such as confidence levels and intersection over union. Yet, in the specific context, it might be preferable to accept some false positives rather than increase accuracy with the potential of missing an enemy sniper.

Figure 8. False positive detections

(a) False positive 1

(b) False positive 2

Discussion

This study aimed to study the feasibility of using a pre-trained YOLOv8 as a platform to develop an effective Pre-Shot Sniper detector. As the gaps of the literature review revealed, for this purpose a high-quality custom Pre-Shot Sniper image dataset was built from scratch.

The experiments have established that a minimum of 350 images from the dataset is sufficient to achieve a mean Average Precision (mAP50) of at least 85% during validation. A larger training set size did not significantly enhance the mAP50. However, it did contribute to a 2% improvement in mAP50-95 during validation, indicating that for critical applications where bounding box accuracy is essential, increasing the training set size might prove advantageous. Further research is required to draw definitive conclusions.

The dataset employed in the study was categorized into several distinct environmental backgrounds: Desert (characterized by sandy, yellow terrain), Snow (encompassing icy, white landscapes), and Rural (including diverse greenery such as forests, tundras, and meadows). Due to a limited number of images, the Urban category, defined by dense cityscapes and buildings, was excluded from the comparative analysis. Examination of these varied backgrounds revealed that it was easiest to spot snipers in a rural environment and hardest to detect them in a snowy background.

In this study, we further evaluated the trade-off between the precision of bounding box annotations and the quantity of images annotated. The findings demonstrate that less precise bounding boxes considerably lower the mAP50-95. Therefore, in scenarios where the accuracy of the bounding box is crucial, it is more beneficial to focus on precise annotations. Alternatively, if the objective leans more towards object detection rather than precise spatial localization, then a less stringent approach to bounding box precision, leading to a larger training set, might be a more effective strategy.

In the process of annotation, certain images were labeled with only the sniper rifle, while others included both the sniper and the rifle. Initially, the intention was to label the sniper alongside the rifle. However, this was not always feasible due to the annotation's size sometimes encompassing almost the entire image. This limitation might have contributed to a reduced mean Average Precision (mAP). Nonetheless, given the limited number of images available for the study, this approach appeared to be the most suitable.

We experienced a lack of images and videos for training and inference of more realistic or real-world scenarios i.e. varying enemy sniper distances and different weather conditions (i.e. heavily snowing, heavily raining, stormy/windy, misty, and twilight). These limitations can affect the Pre-Shot Sniper detector performance.

In conclusion, although some limitations exist, the YOLOv8 as a platform contributes effectively to the design, speed, and accuracy of detecting Pre-Shot snipers. Equipped with a high-end camera sensor, soldiers' capabilities of detecting enemy snipers early can be further improved. Further research should be focused on addressing features of different and/or extreme real-world weather conditions using this special camera sensor, including studying the effects of higher resolution images, especially concerning distance limitations. The Pre-Shot sniper detector (should) act as an extra pair of eyes for soldiers, continuously scanning for (hidden) enemy snipers under different situations and conditions. We encourage DOD to use this YOLO platform and to use the Pre-shot dataset and the Pre-Shot detector as a base for further accuracy improvements.

Access to the codebase used in this report is provided on the GitHub page. Similarly, the dataset crafted for this investigation is hosted and available for access on Roboflow.

References

ANOVA test. (2024). `https://en.wikipedia.org/wiki/Analysis_of_variance`.

Barişik, E., & Baltacıoğlu, G. (2014). The employment of sniper in modern battlefield. *Journal of Management and Information Science*, *2*(1), 13–19.

Bartlett's test. (2024). `https://en.wikipedia.org/wiki/Bartlett%27s_test`.

BBC news. (2023). `https://https://www.bbc.com/news/world-europe-66354363`.

Bhatti, M. T., Khan, M. G., Aslam, M., & Fiaz, M. J. (2021). Weapon detection in real-time cctv videos using deep learning. *IEEE Access*, *9*, 34366–34382.

Murugan, P., Merry ida, A., Aashika, R., Amshiga Brillian, S., Seals Ancy, M., & Sneha, S. (2023). Weapon detection system using deep learning. *https://ijirt.org/master/publishedpaper/IJIRT160929_PAPER.pdf*.

Narejo, S., Pandey, B., Esenarro Vargas, D., Rodriguez, C., & Anjum, M. R. (2021). Weapon detection using yolo v3 for smart surveillance system. *Mathematical Problems in Engineering*, *2021*, 1–9.

NBC news. (2006). `https://www.nbcnews.com/id/wbna15336981`.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 779–788).

Shapiro-Wilk test. (2024). `https://en.wikipedia.org/wiki/Shapiro-Wilk_test`.

Sniper. (2024). `https://en.wikipedia.org/wiki/Sniper`.

Terven, J., & Cordova-Esparza, D. (2023). *A comprehensive review of yolo: From yolov1 and beyond.*

Ultralytics Performance Metrics. (2023). `https://docs.ultralytics.com/guides/yolo-performance-metrics/`.

US Ministry of Defense Patent. (2015). `https://patents.google.com/patent/US9488442B2/en`.

Yadav, P., Gupta, N., & Sharma, P. K. (2023). A comprehensive study towards high-level approaches for weapon detection using classical machine learning and deep learning methods. *Expert Systems with Applications*, *212*, 118698.