Real Time Object Detection for Competitive Robotics

Jordan Chung
Queen's University
jordan.chung@queensu.ca

Andrew Gault
Queen's University
22xjs5@queensu.ca

Ela Aydiner
Queen's University
23pl48@queensu.ca

Wafeeqa Chowdhury
Queen's University
wafeeqa.c@queensu.ca

Daniel Quinn
Queen's University
20dtq@queensu.ca

Armaan Singla
Queen's University
22xv36@queensu.ca

Abstract—The real-time detection of objects in competitive robotics, particularly for competitions such as RoboMaster, is critical for rapid and precise decision-making. This study focuses on developing a robust object detection model utilizing YOLOv5, optimized for identifying opponent robots' armor plates in real-time. The model was trained using publicly available RoboMaster datasets and implemented data augmentation techniques to enhance its generalization capabilities. Evaluation metrics including precision, recall, and mean Average Precision (mAP) demonstrated strong overall performance, achieving 95.1% precision, 97.2% recall, and 98.7% mAP at an IoU threshold of 50%. Despite impressive performance at moderate thresholds, stricter IoU criteria showed lower mAP scores, highlighting areas for future improvements. Ethical considerations, including privacy, transparency, and fairness, were also addressed. The advancements in this object detection model have broader implications, notably in emergency response and healthcare, signifying its potential cross-industry impact.

I. Introduction

A. Motivation

The field of competitive robotics presents an environment where precision and rapid response are not merely advantageous but essential. In competitions such as RoboMaster, robots must navigate through complex arenas filled with obstacles while simultaneously tracking opponent robots, requiring real-time, precise object detection and tracking to allow for split second decisions. This high-pressure environment serves as an ideal testbed for developing advanced AI models that combine rapid response with meticulous precision. By honing these algorithms under competitive conditions, research is not only enhancing the performance of robotic systems in tournaments but are also generating valuable insights into sensor fusion, real-time data processing, and neural network optimization that are transferrable to a broad spectrum of applications.

Our motivation for this project reaches beyond the competitive arena, also to the technological advances derived from object detection research have far-reaching implications in other critical sectors, with the potential to save countless lives. One prominent application is in emergency search and rescue response, where unmanned aerial vehicles (UAVs), provide numerous advantages to rescue operations. Object detection models, similar to those used in robotics competitions, have the potential to significantly boost the efficiency of these operations, which often require a lot of man power [Abbas et al., 2024]. This capability can dramatically reduce search times, support in disaster assessment and overall improve the efficiency and safety of rescue operations, potentially saving countless lives during critical moments.

Healthcare offers another compelling domain where these object detection technologies can be transformative. In the detection of tumors, for example, object detection poses a promising solution. With brain tumors in particular, early detection is crucial, as they can spread throughout the brain at a fast rate [Boesch, 2023]. Object detection applications, can aide medical professionals, making the detection of tumors faster, and reducing human error, potentially saving the lives of many patients [Boesch, 2023]. Such innovations underscore the profound impact that refined object detection technologies can have on both patient safety and the overall efficiency of healthcare delivery.

By addressing the dual challenges of high-speed competitive robotics and critical cross-industry applications such as emergency response and healthcare, this research effort embodies both principles of theoretical innovation and practical impact. The competitive robotics environment acts as a crucible, refining AI models under conditions of extreme speed and precision, while the lessons learned directly inform and enhance applications that carry significant societal benefits. Ultimately, the development of robust, real-time object detection systems not only paves the way for advancements in robotic competitions but also holds the promise of revolutionizing sectors where precision can have life-altering consequences. This interdisciplinary approach highlights the transformative potential of AI-driven object detection, positioning it at the forefront of both technological innovation and practical, real-world application.

B. Problem Definition

The paper will focus on the development of an object detection model which can accurately and precisely detect

the location of an opponent robot's armour plate in real-time. More specifically, this paper will explore the YOLOv5 model by ultralytics [ult,]. The model will be run on-device using the NVIDIA Jetson JetPack and will be trained on online data from past RoboMaster competitions.

II. RELATED WORK

Recent advancements in real-time object detection have significantly improved UAV-based applications, particularly in emergency search and rescue operations, where unmanned aerial vehicles (UAVs) play a crucial role in quickly identifying people, obstacles, and hazards. Researchers Wu et al. introduced YOLOv5_mamba, an optimized YOLOv5 model designed to enhance small-object detection in aerial imagery [Wu et al., 2024]. Their approach refines YOLOv5's backbone by integrating the C2f module, which improves how the model extracts and retains important image details, particularly for small and hard-to-see objects [Wu et al., 2024]. They also incorporate a bidirectional dense feedback network (BDFN), which allows different parts of the model to exchange information across multiple layers, helping it recognize objects more accurately even in challenging conditions like motion blur or partial occlusion [Wu et al., 2024]. Finally, an adaptive gate feature fusion mechanism is introduced to help the model prioritize the most relevant details in an image while filtering out unnecessary noise, making detections more precise [Wu et al., 2024].

Wu et al. tested these improvements on the VisDrone2019 dataset, achieving a 9.3% improvement in mean average precision (mAP) compared to the standard YOLOv5 model [Wu et al., 2024]. Their modifications address common UAV detection challenges, such as viewpoint variations, motion blur, and detecting objects in low-resolution images. By improving how features are processed and shared throughout the model, YOLOv5_mamba significantly enhances small-object detection, making it highly applicable to real-world search and rescue missions.

Although this work is focused on UAV-based detection, many of the technical challenges it addresses—such as detecting small, fast-moving, and partially hidden objects in dynamic environments—are directly relevant to real-time robotic vision in competitive settings like RoboMaster. UAV-based emergency response relies on fast, accurate detection models to locate people in disaster zones, just as robotic competitions require rapid identification of opponent armor plates while dealing with movement and occlusions. The feature-sharing techniques and filtering mechanisms introduced by Wu et al. could be adapted to robotic applications, where rapid changes in the environment make object detection difficult [Wu et al., 2024]. By applying UAV-inspired improvements to YOLOv5, our research explores how these optimizations can enhance object detection in high-speed robotics, strengthening the connection between autonomous aerial surveillance and competitive robotic vision.

Beyond UAV-based applications, real-time object detection has also been explored in medical imaging, where precise identification of abnormalities is crucial for diagnosis and treatment. Aldughayfiq et al. developed a YOLOv5-based deep learning model for pressure ulcer detection, focusing on the early identification and classification of ulcers in patients with limited mobility [Aldughayfiq et al., 2023]. Pressure ulcers, also known as bedsores, form when prolonged pressure on the skin restricts blood flow, leading to tissue damage. If left untreated, these ulcers can worsen, increasing the risk of infection and serious medical complications. Early detection is essential for preventing severe cases and improving patient outcomes. To improve detection accuracy, Aldughayfiq et al. trained a YOLOv5 model to classify ulcers into four severity stages, using data augmentation and transfer learning to make the model more reliable across different patients [Aldughayfiq et al., 2023]. Data augmentation involved artificially expanding the dataset by applying small modifications to existing images—such as rotating, flipping, or adjusting brightness—to help the model generalize better. Transfer learning allowed them to start with a pre-trained YOLO model (originally trained on large-scale datasets) and fine-tune it specifically for ulcer detection, reducing training time and improving accuracy [Aldughayfiq et al., 2023]. Additionally, they optimized multiscale anchor boxes, which help the model detect ulcers of different sizes more effectively, a technique also used in small-object detection tasks like UAV-based imaging [Aldughayfiq et al., 2023]. As a result of these optimizations, their model achieved an overall mean average precision (mAP) of 76.9%, significantly improving detection performance compared to traditional methods [Aldughayfiq et al., 2023]. This study highlights how real-time object detection models can assist healthcare professionals by providing an automated system for early ulcer detection, reducing human error and improving patient care. Although this research is centered on medical image analysis, the core machine learning challenges it addresses—such as recognizing small features, refining feature extraction, and ensuring real-time classification—are directly relevant to real-time armor detection in RoboMaster competitions. Just as YOLOv5 is trained to detect subtle patterns and variations in ulcer severity, similar techniques can be applied to detecting small, partially occluded armor plates on fast-moving robots. The use of multi-scale anchor boxes, adaptive feature selection, and transfer learning in ulcer detection suggests strategies that could help enhance real-time robotic tracking and classification. By integrating advancements from both UAV-based object detection and AI-driven medical imaging, our research explores how YOLOv5 optimizations can improve real-time object detection in competitive robotics. The ability to detect critical targets under dynamic conditions—whether in emergency response, healthcare, or robotics—highlights the broader impact of object detection advancements across multiple fields.

III. METHODOLOGY

A. Dataset

For our project, we used the RoboMasters dataset available on RoboFlow, which contains 2,779 images. This dataset was

sourced online and provided a foundation for training our AI model. However, it is important to note that the Queen's Knights Robotics Team (QKRT) has its own dataset, which we did not have access to. The dataset we used primarily consists of images related to the RoboMaster competition, featuring various robots and environments relevant to our application. While the dataset offered sufficient diversity in terms of lighting conditions and robot positions, data augmentation techniques were employed to artificially expand the dataset and improve model generalization. 40% of the dataset was augmented by randomly rotating the image +/- 8 degrees. These augmentations helped the model learn more robust feature representations, reducing the risk of overfitting to the limited dataset.

B. Preprocessing

To expand the dataset and diversify the data, we preprocessed our training data by modifying every image, creating several adjusted versions of the original images, all combined together into a much larger dataset. First, we rotated every image randomly between $-8°$ and $8°$, not too large of a maximum rotation so as to minimize blank space. In order to maintain the validity of the labels, they, too, had to be rotated the same way. Each label was converted into its absolute value counterpart, the corners calculated, then the rotation matrix was applied to each corner of the label, then converted back into YOLOv5 format. We then increased the contrast of every image by an alpha value of 1.5. Finally, to simulate artifacts and potential noise created by the camera, a layer of translucent coloured random noise was added to every image. This all created four sets of the original dataset: The original, rotated, contrast adjusted, and noise versions, which were all combined into the dataset the model was trained on.

C. Modelling

YOLOv5, developed by Ultralytics, is incredibly efficient, has real-time processing capabilities, and creates a balance between speed and accuracy. The model is implemented in PyTorch, making it accessible for both training and deployment, while offering multiple variants (YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x) that allow for further flexibility depending on constraints and the project itself. Next, its anchor-based detection method assures reliable object localization, which is important for our application. Then, compared to other YOLO versions, YOLOv5 provides an optimized balance between model size, speed, and detection accuracy, making it a solid model to use. Additionally, its robust and generalizable architecture, along with its refined processing and data augmentation, allows for adaptation to many environments. So, given our focus on real-time object detection, particularly in autonomous systems, YOLOv5's fast inference speed and refined detection capabilities make it the ideal choice for our implementation.

D. Evaluation Methods

In evaluating the performance of the YOLOv5 model, mean Average Precision (mAP), Precision, and Recall were used as the primary indicators of detection quality. The mean Average Precision (mAP) offers a holistic measure of how accurately the model detects and localizes objects across various Intersection over Union (IoU) thresholds. In our experiments, we tracked both mAP@0.5 (which uses a fixed IoU threshold of 0.5) and mAP@0.5:0.95 (which averages performance over multiple IoU thresholds from 0.5 to 0.95). An upward trend in these values indicates that the model progressively refines its bounding box predictions. The higher the mAP, the more reliably the system distinguishes between true positives and negatives, and the better it captures precise object boundaries. .
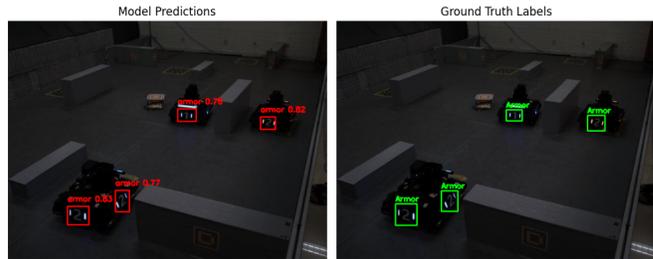
IV. RESULTS AND DISCUSSION

Fig. 1. Example Model Output Compared with Ground Truth

TABLE I
SUMMARY OF THE BENCHMARKS PRODUCED IN THE TESTING OF THE TRAINED YOLOV5 MODEL.

Precision	Recall	mAP 50	mAP 95
95.1%	97.2%	98.7%	52.1%

A. Analysis

Upon analysis of the obtained results, clear insights into the model's performance and behavior are evident. The evaluated metrics demonstrate a strong performance overall, particularly highlighted by high precision and recall values. Specifically, the precision value achieved was 0.951, indicating that approximately 95.1% of bounding boxes predicted by the model corresponded accurately to actual objects, suggesting minimal false positives. Additionally, the recall score of 0.972 signifies that the model successfully detected about 97.2% of the actual instances, demonstrating its strong capability in identifying the majority of objects, resulting in few false negatives.

Further examining the model's capability, the Mean Average Precision at IoU threshold of 50% (mAP50) yielded an impressive score of 0.987 (98.7%). This high value reflects excellent performance when employing a moderate overlap threshold, indicating that the model excels in general object detection scenarios. However, when stricter localization criteria were considered—specifically, the mean average precision computed across IoU thresholds ranging from 50% to 95% (mAP50-95)—the model's performance decreased to 0.521 (52.1%). This drop is typical as the IoU thresholds become

stricter, demanding more precise bounding box localization. Thus, while the model is robust in general detection tasks, there remains room for improvement in scenarios requiring precise localization.

Regarding the efficiency of the model, the speed metrics are satisfactory for practical deployment scenarios. Image pre-processing averaged 3.6ms per image, indicating quick preparation of input data through resizing and normalization. Inference time, the most significant portion, averaged at 338.5ms per image, which is acceptable but highlights potential areas for optimization, especially for real-time applications. Finally, Non-Maximum Suppression (NMS), critical for refining detection outputs by filtering overlapping boxes, only required 1.4ms per image, emphasizing its minimal impact on the total processing time.

Overall, these results highlight the model's effectiveness and pinpoint specific aspects—particularly precise localization at stricter thresholds and inference time optimization—that could benefit from further refinement. Future efforts could involve fine-tuning model parameters or exploring advanced architectures to address these identified areas, ultimately enhancing performance in applications demanding high accuracy and efficiency.

B. Ethical Considerations

When designing and implementing a robot detection model, managing the ethical considerations of privacy, transparency, and fairness is crucial. Privacy is a primary concern, especially if the model is used in a team setting, as it affects how training data is collected, used, and stored. This may include sensitive data, such as real-time images of robots and people, which should be securely managed if stored and have clear data retention policies outlining how long it will be stored before being deleted. It is essential to ensure that only the necessary data for the intended purposes are collected. Next, transparency is just as important, requiring thorough documentation that explains how the model processes information and makes decisions in case of an error. Fairness is also a key ethical concern, as there may be time and resource disparities between teams, which can influence model performance or prevent them from using the technology altogether. Therefore, in competitive settings, making these solutions open source and accessible could help even the playing field. Additional technological solutions, such as camera control that limits the capture of sensitive information or anonymization techniques, can further improve ethical precautions. Addressing these considerations will help ensure that the model is developed responsibly and used in a way that aligns with ethical best practices.

V. Future Work

In future works, we plan to make our model more suitable for competition use by training it on specialized images created by our client, QKRT, tailored explicitly to their use-case scenarios. Additionally, we will evaluate the model's inference performance and suitability for real-time deployment on NVIDIA's Jetson JetPack platform, aiming to achieve optimal speed and efficiency in practical, resource-constrained environments.

VI. Limitations

Despite implementing pre-processing and augmentation techniques, our project faced several key limitations. One of the primary challenges was the size of the dataset. With a limited number of images, the dataset was relatively small, limiting the effective training of a computer vision model, leading to potential generalization issues. The lack of access to the QKRT's internal dataset further constrained our ability to train our model on data that is specific to our team's dataset.

Hardware limitations also impacted our ability to develop and fine-tune the model. Due to computational constraints, there were difficulties training the model on the available hardware. This significantly slowed experimentation and iterative improvements.

VII. Conclusion

This research demonstrated the effectiveness of the YOLOv5 model in real-time object detection scenarios critical to competitive robotics, achieving high precision and recall rates. Specifically, the model obtained a precision of 95.1%, a recall of 97.2%, and a mean Average Precision (mAP) of 98.7% at a 50% IoU threshold. However, performance decreased under stricter IoU thresholds, indicating room for enhanced localization precision. Future work will include training the model on specialized datasets from Queen's Knights Robotics Team (QKRT) to further optimize its accuracy and speed for real-time deployment. Additionally, addressing hardware and dataset limitations will help improve the model's robustness and generalization capabilities. These enhancements not only promise improvements in robotic competitions but also demonstrate potential transformative impacts across other critical fields, such as emergency rescue operations and medical diagnostics.

References

[ult,] Yolov5 documentation. https://docs.ultralytics.com/yolov5/.

[Abbas et al., 2024] Abbas, Y., Mudawi, N. A., Alabdullah, B., Sadiq, T., Algarni, A., Rahman, H., and Jalal, A. (2024). Unmanned aerial vehicles for human detection and recognition using neural-network model. *Frontiers*.

[Aldughayfiq et al., 2023] Aldughayfiq, B., Ashfaq, F., Jhanjhi, N., and Humayun, M. (2023). Yolo-based deep learning model for pressure ulcer detection and classification. *Healthcare (Basel)*, 11(9):1222.

[Boesch, 2023] Boesch, G. (2023). Top 19 applications of computer vision in healthcare. https://viso.ai/applications/computer-vision-in-healthcare/.

[Wu et al., 2024] Wu, S., Lu, X., and Guo, C. (2024). Yolov5$_{m}amba$: $unmannedaerialvehicleobjectdetectionbasedonbidirectionaldensefeedbacknet$