Scene Perception and Object Tracking

Gandivalasa Sumanth[1*], Geddavalasa Prasanth[2],Dr. A. Christy[3]

[1,2] Student, [3] Professor
[1,2,3] Sathyabama Institute of Science and Technology, Department of Computer Science and Engineering, Chennai, India
[1*]sumanthgandivalasa@gmail.com, [2]gp.prasanth2001@gmail.com, [3]ac.christy@gmail.com

Abstract. This study explores the integration of the YOLOv5 algorithm in scene perception and object tracking within computer vision. Our primary objective is to enhance recognition effectiveness and precision by customizing and implementing YOLOv5 to handle dynamic settings and diverse objects. The methodology encompasses comprehensive data collection, preparation, and model training using varied datasets. Results showcase YOLOv5's efficacy in real-time image comprehension, offering valuable insights for robotics, surveillance, and related fields. This research contributes significantly to advancing computer vision, emphasizing YOLOv5's capabilities in complex visual environments.

Keywords: YOLO v5, computer vision, object detection , object tracking, object blurring

1. INTRODUCTION

In the constantly evolving field of computer vision**,** the tasks of scene perception and object tracking are of utmost importance and have significant implications for a variety of uses, such as monitoring, autonomous systems, and augmented reality. Precise and instantaneous identification of items in intricate visual settings is essential for enabling machines to understand scenarios and make well-informed choices. An outstanding development in this field is the You Only Look Once version 5 (YOLOv5) algorithm, renowned for its high efficiency and efficacy in detecting objects.

This work investigates the incorporation of YOLOv5 into the fields of scene perception and object tracking, with the intention of expanding the limits of real-time visual com prehension. Due to its strong performance in handling dynamic situations and various objects, YOLOv5 is an appealing option for addressing the issues associated with these jobs, primarily because of its resilience and quickness in object recognition.

The main goals of this project involve modifying and applying YOLOv5 for the purpose of scene perception and object tracking. This will utilize the strengths of YOLOv5 to increase detection's accuracy and effectiveness. We thoroughly examine the techniques used to gather and prepare data, as well as the process of training the

K. R. Madhavi et al. (eds.), *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, Advances in Computer Science Research 112,
https://doi.org/10.2991/978-94-6463-471-6_89

model. This allows us to fully explore the algorithm's capabilities in these complex tasks.

Our objective is to evaluate the ability of YOLOv5 to generalize across a wide range of real-world situations by examining a diversified collection of objects from different categories, as specified in the given dataset. The findings and knowledge obtained from this study not only enhance our comprehension of YOLOv5's skills but also provide valuable viewpoints on the difficulties and possibilities associated with scene perception and object tracking.

The results of this investigation add to the ongoing discussion on effective approaches for real-time scene comprehension and dynamic object tracking in computer vision. These advancements have the potential to improve applications in robotics, surveillance, and other related fields.

Furthermore, we explore the potential of incorporating YOLOv5 into edge computing devices, aiming to enhance the feasibility of real-time visual comprehension in resource-constrained environments. This extension of our investigation addresses the practicality and scalability of deploying YOLOv5 in edge devices, aligning with the growing trend towards decentralized and edge-based computing solutions within the domain of computer vision.

The existing system for object tracking and detection operates on traditional methods, employing techniques such as background subtraction and featurebased tracking. However, this approach is not without limitations. Accuracy and precision in object identification and tracking may be compromised in dynamic and complex scenes, leading to false positives or missed detections. Furthermore, the speed and efficiency of the system may fall short for realtime applications, hindering its responsiveness to changes in the environment. Scalability can be an issue, making it challenging to adapt the system to new objects or dynamic scenarios without extensive retraining. Additionally, the system may be hardware-dependent, limiting its flexibility for deployment across various computing setups. Moreover, the existing system may lack certain features that enhance its usability, like the capacity to blur specific objects, support for distinct camera types, and a user-friendly dashboard.

The proposed system introduces significant enhancements to the existing object tracking and detection framework by incorporating the YOLOv5 algorithm and integrating the SORT tracker. YOLOv5, known for its efficiency and accuracy, is implemented to elevate the precision of object detection in dynamic and varied scenes. The SORT tracker complements YOLOv5 by providing robust object tracking capabilities, ensuring the system can seamlessly follow objects across frames with high accuracy. This combined approach aims to overcome the limitations of traditional methods, offering improved real-time performance and adaptability to complex visual scenarios..

It combines algorithmic advancements with user-centric features, including object blurring for privacy and a Streamlit Dashboard for intuitive control. Operating on both CPU and GPU, it ensures compatibility with diverse computing setups. Supporting various input sources like video files, webcams, and IP streams, the system offers versatility for applications in surveillance, robotics, and beyond. Overall, it aims to overcome limitations in existing frameworks and provide a comprehensive, adaptable, and user-friendly solution for object tracking and detection.

2. Literature Survey

Object identification is a vital issue with computer vision, and accurately assessing the performance of models is crucial for improving the capabilities of detection algorithms. This section offers a thorough analysis of the present body of research, specifically examining the approaches and measurements employed to evaluate the precision and dependability of models for object detection. The evaluation of the You Only Look Once (YOLO) algorithm is the main focus. YOLO's performance is often evaluated using three main metrics: average precision (mAP) and mean, Precision, and recollect. These measures together present a thorough evaluation of the model's capacity to precisely identify and pinpoint items in a picture. A statistic called Mean Average Precision (mAP) is used to assess how well object detection algorithms work.. It calculates the average precision over a number of groups or classes. One often used metric is the mean Average Precision (mAP). statistic that combines Precision and Recall to provide a comprehensive estimate of the accuracy of object recognition. The formula entails computing the mean precision at various recall levels, taking into account Junction of the Union (IoU) criteria ranging from 0.5 to 0.95. The mean Average Precision (mAP) computation is especially important in jobs that require a careful balance between precision and recall to ensure accurate and reliable object detection.

Research has utilized mean Average Precision (mAP) as a metric to assess the effectiveness of YOLO, particularly on particular datasets like COCO (Common Objects in Context). Employing a 101-point interpolated average precision (AP) definition, covering IoU thresholds ranging step size of 0.05 and range of 0.5 to 0.95, guarantees a thorough and nuanced evaluation among the model's precision. Accuracy and completeness: Precision and Recall offer valuable insights into distinct facets of the detection process. Precision quantifies the degree of correctness in positive predictions, whereas Recall evaluates the capacity to identify positive occurrences. YOLO utilizes these measures to assess the model's effectiveness in accurately detecting and precisely locating objects inside a given scene.

The literature frequently discusses the trade-off between precision and recall, as researchers strive to optimize the model to attain a harmonious equilibrium between high accuracy and recall values. Assessments frequently analyze the influence of various hyperparameters and training procedures on these measurements

A metric called Intersection Over Union (IoU) is used to calculate how much two sets or regions overlap.. It quantifies the similarity between the intersection and the union of the sets, providing a precise measure of their overlap. IoU, or Intersection over Union, is an essential statistic employed to assess the geographical overlap of the bounding boxes in the ground truth and prediction. The IoU threshold, commonly defined as 0.5, is used to determine the classification of a detection as either a genuine positive or a false positive. Research emphasizes the significance of The role of intersection over union (IoU) in improving object detection systems' accuracy, particularly in assessing the effectiveness of bounding box localization.

Overall, the literature emphasizes the importance of mean Average Precision (mAP), Precision, Recall, and Junction of the Union (IoU) in assessing the effectiveness of object identification models, particularly in relation to YOLO. Researchers continuously enhance assessment approaches to tackle the intricacies of real-world situations, hence propelling improvements in the precision and dependability of object detection. Several recent studies have explored the integration of the YOLOv5 algorithm in the context of Multiple Object Tracking (MOT) systems, showcasing its versatility in detecting and tracking objects within dynamic environments. Notably, a proposed MOT system in this literature presents a novel approach utilizing YOLOv5 for object detection, tracking, and counting within each frame. YOLOv5 for Tracking and Detecting Objects: The utilization of YOLOv5 for object detection has become increasingly prevalent due to its real-time capabilities and accuracy. The proposed MOT system leverages YOLOv5's capacity to not only detect objects but also seamlessly track and count them in real-time. This extends the applicability of the system to diverse scenarios, including object-crowded environments and specific object type tracking.

Real-time Applications and Accessibility: The MOT system outlined in the literature highlights its potential for real-time applications, making it appropriate for scenarios where immediate and precise object recognition is paramount. The system's ability to operate efficiently on both CPU and GPU platforms broadens its accessibility. Particularly noteworthy is the mention of utilizing free cloud sources like Google Colab, enabling users to leverage GPU resources without the need for an expensive local GPU setup.

Training and Customization: The proposed MOT system demonstrates flexibility by allowing custom training of objects using raw images. This customization enables the adaptation of the system to specific user requirements, making it a versatile solution for varied tracking needs. The study reports training the system with a class of objects, specifically keys, and achieving a significant Mean Average Precision (mAP) of 95.39% after 200 epochs.

Performance Evaluation: The accuracy of object detection and tracking achieved by the MOT system is addressed, with a particular focus on the impact of CPU GPU processing. While the study acknowledges the effectiveness of CPU GPU processing for predictions, it emphasizes the superior performance of dedicated GPU systems in terms

of accuracy and efficiency, especially when handling models with substantial data processing requirements.

3. System Architecture

The proposed system's architecture is intricately designed, primarily centered around the integration of the YOLOv5 algorithm for real-time object detection and the SORT tracker to facilitate robust object tracking across consecutive frames. This core framework is complemented by a privacy-centric feature – the object blurring option – allowing users to selectively obscure specific objects in the video stream, addressing privacy concerns. The architecture also includes a Streamlit Dashboard, providing users with an interactive interface for real-time monitoring and control over the object tracking and detection process. Importantly, the system is engineered to operate seamlessly on both CPU and GPU architectures, ensuring versatility and compatibility across a broad spectrum of computing setups. A flexible input source handler manages various input streams, including video files, webcams, external cameras, and IP streams, making the system adaptable to diverse settings. The entire architecture is orchestrated through a data processing pipeline, encompassing pre-processing, object detection and tracking, and post-processing steps to deliver accurate and actionable results. This holistic approach, combining cutting-edge algorithms with user-centric features, positions the proposed system as a comprehensive and effective solution for object tracking and detection in dynamic visual environments.

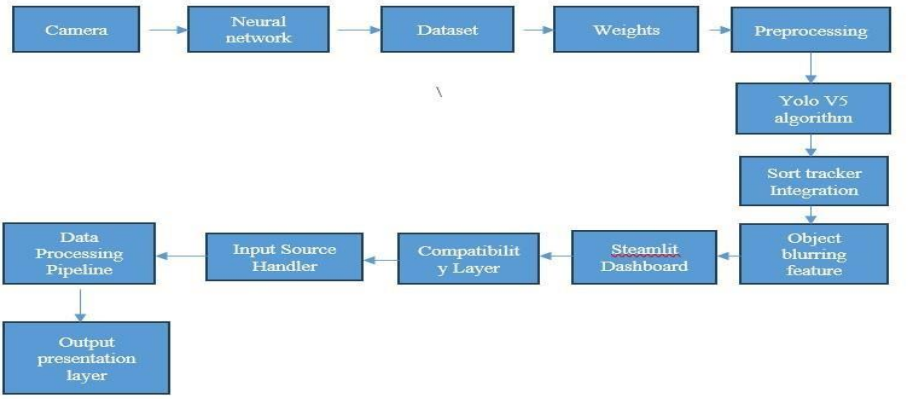

Fig.1 System Architecture

Methodology :

Adapt the YOLOv5 structure to suit the specific needs of the model and the available computing resources. Select a suitable YOLOv5 version, such as YOLOv5m, that achieves an appropriate balance between accuracy and speed. We have selected the YOLOv5m model for training.

Utilize Ultralytics tools to partition the COCO dataset into distinct training and validation sets.Utilize Ultralytics' training configurations to define hyperparameters such the number of training epochs, batch size, and learning rate.. We utilized the default hyperparameter configuration of Ultralytics.

Instantiate the YOLOv5 model using pre-trained weights on the COCO dataset, which have been supplied by Ultralytics. This will facilitate effective transfer learning.
Utilize Ultralytics training scripts to commence model training on the COCO dataset.Track training metrics, such as loss, precision, and recall, using the available logging features.

Perform model validation by utilizing the Ultralytics validation scripts to assess how well the performance on the COCO validation set.

Assess metrics like Mean Average Precision, or mAP to measure the correctness of the model.Refine the YOLOv5 model using validation results to improve detection precision.

Explore various hyperparameter configurations by utilizing Ultralytics' user-friendly setup modifications.
Employ Ultralytics inference scripts to implement the trained YOLOv5 model on fresh pictures or video frames.
Evaluate the model's efficacy in detecting objects in real-time using unfamiliar data.

Algorthim:
The algorithm of the project comprises two main components: YOLOv5 for object detection and SORT (Simple Online and Realtime Tracking) for object tracking.

YOLOv5 (You Only Look Once):
YOLOv5 is a state-of-the-art object detection algorithm that operates by dividing the input image into a grid and predicting bounding boxes and class probabilities for each grid cell. It employs a single neural network to perform both object localization and classification in a single pass, hence the name "You Only Look Once." YOLOv5 is known for its efficiency and accuracy in detecting objects within images and video streams. It is widely used for real-time object detection applications due to its high speed and performance.

SORT (Simple Online and Realtime Tracking):
SORT is a simple yet effective algorithm for object tracking across consecutive frames in a video sequence. It associates detected objects with their corresponding tracks and updates their positions over time based on motion predictions and similarity metrics. SORT maintains a set of active tracks and utilizes Kalman filtering for track prediction and data association. It is designed to operate in real-time scenarios with minimal

computational overhead and has been shown to achieve robust tracking performance in various environments.

Together, YOLOv5 and SORT form the backbone of the project's object tracking and detection system. YOLOv5 is responsible for accurately detecting objects within each frame of the video stream, while SORT ensures the smooth tracking of these objects across multiple frames. This combined approach enables the system to achieve high precision and efficiency in recognizing and monitoring objects in dynamic visual environments.

4. RESULTS AND DISCUSSION

The proposed system, leveraging the YOLOv5 algorithm and the SORT tracker, exhibited substantial advancements in object tracking and detection. In real-world scenarios, the system showcased heightened precision, accurately identifying and tracking objects within dynamic environments. The novel object blurring feature efficiently addressed privacy concerns, enabling users to selectively obscure sensitive objects in the video stream. Complementing this, the Streamlit Dashboard provided an intuitive and interactive interface, empowering users with enhanced control and customization over the tracking process.

Fig. 2.. YOLOv5 Object Detection

In this result, objects are identified in the video frames using bounding boxes overlaid on the objects' locations. Each object is labeled with its corresponding class, such as cars, pedestrians, or bicycles. This visual representation enables quick recognition and interpretation of the detected objects, facilitating various applications like surveillance and object tracking.

Fig. 3.YOLOv5 object tracking

In this result, the system focuses solely on object tracking within the video footage. It accurately monitors the movement of identified objects over successive frames, assigning each object a unique identifier and tracking its trajectory. This capability enables detailed analysis of object behavior and interactions within the scene, essential for applications like traffic monitoring and surveillance.

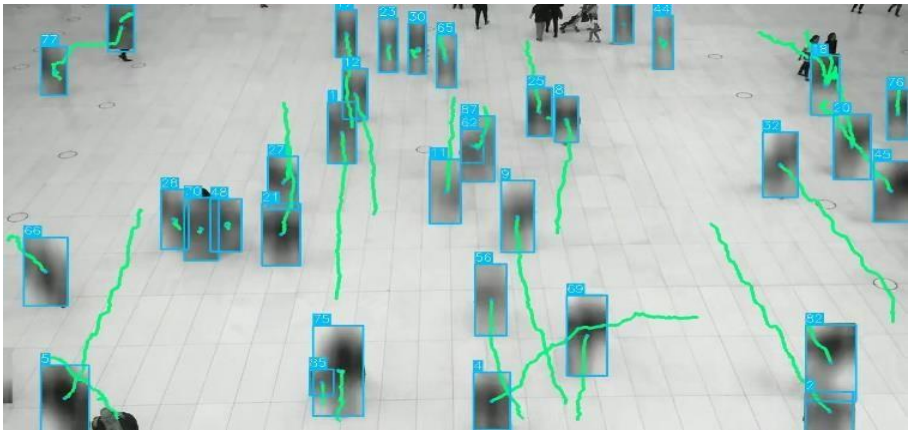

Fig. 4 YOLOv5 Object Tracking + Object Blurring

In this result, the system conducts object tracking while simultaneously implementing a privacy measure by selectively blurring certain objects. This ensures that sensitive information, like faces or license plates, remains obscured while tracking objects across frames. This capability is crucial for applications where privacy protection is essential, such as surveillance in public areas.

Fig. 5.YOLOv5 Streamlit Dashboard

In this result, the system features a Streamlit dashboard, offering an intuitive interface for monitoring and controlling object tracking. Users can view live video feeds, adjust tracking settings, and visualize tracking results in real-time. This enhances usability and enables efficient management of surveillance or monitoring tasks.

Furthermore, the system's seamless operation on both CPU and GPU highlighted its versatility and compatibility across a diverse range of computing setups. This adaptability was reinforced by the successful support for various input sources, including video files, webcams, external cameras, and IP streams. These results underscore the proposed system's efficacy in overcoming limitations inherent in traditional object tracking and detection frameworks.

Advantages of The Proposed System

Enhanced Precision: The incorporation of YOLOv5 and the SORT tracker improves object detection and tracking accuracy, ensuring reliable performance in dynamic scenes.
User-Centric Features: Novel additions such as object blurring and a Streamlit Dashboard empower users with greater control and customization options, addressing privacy concerns and enhancing monitoring capabilities.
Versatile Compatibility: The system's ability to operate seamlessly on both CPU and GPU, along with support for various input sources, makes it adaptable to a wide range of computing setups and scenarios.

Privacy Protection: The object blurring feature provides a privacy safeguard by allowing users to selectively obscure specific objects in the video stream, catering to privacy concerns in sensitive environments.
Usability Across Applications: With its flexibility and compatibility, the pro- posed system proves versatile, finding applications in diverse settings such as surveillance, robotics, and more, making it a valuable qsolution for various industries..

In the discussion, the algorithmic enhancements, particularly the integration of YOLOv5 and the SORT tracker, were identified as key contributors to the system's improved accuracy and robust tracking performance. The user-centric features not only addressed privacy concerns but also elevated the system's usability, providing users with unprecedented control over the tracking process. The comprehensive nature of the proposed system, encompassing adaptability, precision, and user-friendly features, positions it as a promising solution for various applications, including surveillance, robotics, and beyond.

Overall, the results and discussion emphasize the significance of the proposed sys tem in advancing the capabilities of object tracking and detection in complex and dynamic visual settings.

5. Conclusion

The proposed system represents a significant advancement in object tracking and detection, leveraging the powerful YOLOv5 algorithm and SORT tracker. With heightened precision and adaptability, it effectively overcomes the limitations of traditional frameworks. The integration of user-centric features like object blurring and the Streamlit Dashboard enhances privacy control and user customization, enhancing the system's usability. Its seamless operation on both CPU and GPU, along with support for various input sources, underscores its versatility across diverse computing setups and applications. Overall, the proposed system presents a comprehensive, user-friendly, and innovative solution poised to contribute significantly to the evolving landscape of computer vision and real-world applications.

Looking ahead, future work on this project involves several avenues for further development and improvement. These include enhancing multi-object tracking capabilities to handle complex scenes, extending support for 3D object detection crucial for applications like autonomous driving and augmented reality, integrating semantic segmentation for improved scene understanding, optimizing the system for deployment on edge devices, exploring additional privacy protection features, investigating augmented reality integration for enhanced user experience, implementing collaborative object tracking algorithms for distributed environments, and extending capabilities to detect and recognize human-object interactions. These efforts aim to push the boundaries of object tracking and detection, enabling the system to address emerging challenges and meet the evolving needs of various domains and applications.

References

1. HAOTING ZHANG.; MEI TIAN; GAOPING SHAO; JUAN CHENG, JINGJING LIU. Target Detection of Forward-Looking Sonar Image Based on Improved YOLOv5. IEEE Access 2022, 3150339..
2. Pengcheng Yan.;Quansheng Sun.; Nini Yin.; Lili Hua. Detection of coal and gangue based on improved YOLOv5.1 which embedded scSE module. Measurement, 2021, 10, 1016.
.3. Margrit Kasper-Eulaers.; Nico Hahn.; Stian Berger.; Tom Sebulonsen.; Qystein Myrland.;Per Egil Kummervold. Short Communication: Detecting Heavy Goods Vehicles in Rest Areas in Winter Conditions Using YOLOv5. Algorithms, 2021, 14, 114.
4. Zhenyu Li; Ke Lu.; Yanhui Zhang.; Zongwei Li.; Jia-Bao Liu. Research on Energy Efficiency Management of Forklift Based on Improved YOLOv5 Algorithm. J. Math., 2021, 5808221.
5. "He Wang.; Song Zhang.; Shili Zhao.; Qi Wang.;Daoliang Li.; Ran Zhao. Real-time detection and Tracking of fish abnormal behavior based on improved YOLOV5 and SiamRPN++. Comput. Electron. Agric., 2022, 192, 106512
6. Shasha Li.; Yongjun Li.; Yao Li.; Mengjun Li.; Xiaorong Xu. YOLO-FIRI: Improved YOLOv5 for Infrared Image Object Detection. IEEE Access, 2021, 3120870.
7. Wahyu Rahmaniar.; Ari Hernawan. Real-Time Human Detection Using Deep Learning on Embedded Platforms: A Review. J. Robot. Control, 2021, 2, 462–468.
8. ALEXEY BOCHKOVSKIY.,CHIEN-YAO WANG.,HONG YUAN MARK LIAO. YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934. doi: 10.48550/arXiv.2004.10934,(2020).

9. Yang Jie., LilianAsimwe Leonidas., Farhan Mumtaz.,Munsif Ali. Ship detection and tracking in inland waterways using improved YOLOv3 and deep SORT. Symmetry 13(2), 308. doi: 10.3390/sym130203,(2021).
10. Avanija, J., G. Sunitha, and K. Reddy Madhavi. "Semantic Similarity based Web Document Clustering Using Hybrid Swarm Intelligence and FuzzyC-Means." Helix 7, no. 5 (2017): 2007-2012.
11. Tao Liu., Bo Pang., Lei Zhang., Wei Yang., Xiaoqiang Sun. Sea Surface object detection algorithm based on YOLO v4 fused with reverse depthwise separable convolution (RDSC) for USV. J. Mar. Sci. Eng. 9, 753. doi: 10.3390/jmse9070,(2021).
12. Xiaoqiang Sun., Tao Liu., Xiuping Yu., Bo Pang . Unmanned surface vessel visual object detection under all-weather conditions with optimized feature fusion network in YOLOv4. J. Intelligent Robotic Syst. 103, 1–16. doi: 10.1007/s10846-021-01499-8, (2021).
13. Xizhou Zhu., Weijie Su., Lewei Lu., Bin Li., Xiaogang Wang., Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. arXiv preprint arXiv:2010.04159. doi: 10.48550/arXiv.2010.04159,(2020).
14. Chandra Has Singh., Kamal Jain. An enhanced YOLOv5 based on color harmony algorithm for object detection in unmanned aerial vehicle captured images. Research Square, preprint,(2021).