Person Detection from CCTV Footage Using TinyYOLOv3

[1]Radhika K M, [2]Shankar M Bakkannavar, [3]Arjun M S, [4]Samarth Bhaskar Bhat

[1]Final year M.Sc student, [2] Associate Professor, [3]Forensic Tutor, [4]Director

[1-3]Department of Forensic Medicine and Toxicology,

[1-3]Kasturba Medical College, Manipal, India

[4] Reverse Engineering Infosec, Pvt Ltd., Bengaluru, India

Abstract- Person detection from Closed-Circuit Television (CCTV) footage is a crucial task in modern surveillance systems, enabling real-time monitoring and ensuring public safety. Deep learning-based object detection models, particularly Tiny YOLOv3, have emerged as a promising solution for efficient and accurate person detection. This research paper presents a detailed investigation into the implementation and performance evaluation of Tiny YOLOv3 for person detection from CCTV footage. The paper discusses the challenges associated with person detection in CCTV footage, including varying lighting conditions, occlusions, and diverse camera perspectives. It then introduces the architecture and principles of Tiny YOLOv3, highlighting its ability to achieve real-time processing while minimizing computational requirements, making it suitable for surveillance applications. The performance of Tiny YOLOv3 is rigorously evaluated against other state-of-the-art person detection methods using standard metrics such as mean Average Precision (mAP) and processing speed. The results demonstrate the model's effectiveness in accurately detecting persons in complex surveillance environments while maintaining low latency. Furthermore, the paper explores the practical implementation of Tiny YOLOv3 for real-time person detection in a CCTV monitoring system. Overall, this research contributes to the advancement of person detection technologies in surveillance applications, ultimately leading to safer and more secure public spaces.

Keywords- CCTV footage, person detection, tinyYOLOv3, ImageAI, Inceptionv3.

I. INTRODUCTION

Person detection is a fundamental task in the field of computer vision, with significant applications in various domains, including surveillance, robotics, and human-computer interaction. Over the years, deep learning has propelled the progress in person detection algorithms, enabling more accurate and efficient solutions. This research paper introduces a novel approach to person detection that combines the capabilities of three distinct deep learning models: TinyYOLOv3, ImageAI, and Inception V3.

The ability to accurately detect and locate persons in images and videos is of paramount importance in ensuring public safety, enhancing security, and enabling intelligent systems to interact with humans effectively. Traditionally, person detection relied on handcrafted features and traditional machine learning methods, but recent advances in deep learning have demonstrated superior performance in various computer vision tasks.

TinyYOLOv3, a lightweight variant of the YOLO family, has gained popularity for its real-time processing capabilities, making it suitable for resource-constrained environments such as embedded systems and edge devices[1]. ImageAI, on the other hand, offers a high-level library built on top of TensorFlow and Keras, providing pre-trained models for easy and rapid deployment. Lastly, Inception V3, a well-known image classification model, can be adapted for person detection through post-processing techniques. In this paper, we propose a multi-model ensemble approach that leverages the strengths of all three models to enhance the accuracy and robustness of person detection. The ensemble system aggregates the outputs from TinyYOLOv3, ImageAI, and Inception V3 and performs a fusion step to obtain the final detection results. By combining the predictions from multiple models, we aim to mitigate false positives and improve overall detection performance.[2]

To validate the effectiveness of the multi-model approach, we conduct experiments on publicly available person detection datasets, adapting them to the input requirements of each model. The evaluation metrics include precision, and inference speed to assess the ensemble system's accuracy and real-time capabilities. The research focuses on scenarios with complex surveillance environments, such as crowded areas, occlusions, and varying lighting conditions, to ensure the practical applicability of the proposed multi-model approach[3].

This research aims to contribute to the advancement of person detection technology by introducing a novel multi-model ensemble approach. By combining the strengths of TinyYOLOv3, ImageAI, and Inception V3, we seek to achieve enhanced accuracy and real-time performance in person detection tasks. The findings of this research will encourage further exploration of ensemble strategies to improve performance across various computer vision applications beyond person detection.

II. LITERATURE SURVEY

A research by Joseph Redmon et.al (2015) suggested a unified YOLO-based object detection system[4]. They merged many object detecting components into a single neural network. One of the most sophisticated real-time object detectors out there, Fast YOLO was proved to be the quickest object detector on the market. Additionally, YOLO generalised effectively to new domains, which made it the best choice for applications that required quick, reliable object identification. The ability of YOLO to generalise to new domains has been better than that of other detectors, according to an experimental comparison with several object detection systems.

A study made on object detection using YOLO by Tausif Diwan et.al (2022)[5], discussed the challenges, datasets and applications of YOLO. The study discussed classification, localization, and segmentation in single and multiple objects image. The problems faced during object detection such as foreground-background imbalance, detection of smaller objects, inaccurate localization during predictions, etc were put forth. Two stage detection algorithms which included generating region of interest and predicting objects in bounding boxes were studied. RCNN and successors as well as YOLOv1, YOLOv2, YOLOv3, YOLOv4, were explored and evaluated. YOLOs performed significantly better in comparison with their counterpart two stage object detectors in terms of detection accuracy and inference time.

An analytical study on object detection using YOLO was done by Dawn Wilson et.al (2022)[6]. The study explained the working of YOLO algorithm as well as the layers in a convolutional neural network. The authors opine that YOLO follows a regression approach and is fast and accurate. The study does not ignore YOLO's expansive future potential.

III. METHODOLOGY

TinyYOLOv3 is a lightweight variant of the YOLO (You Only Look Once) object detection model, designed to achieve real-time processing and efficient inference on resource-constrained devices such as embedded systems, mobile devices, and edge devices. The model's architecture is based on a single deep neural network and operates on the principle of detecting objects directly from the input image, without requiring any additional proposal generation or post-processing steps. Here's an overview of the working of TinyYOLOv3:

Input Image and Preprocessing:

TinyYOLOv3 takes an input image of fixed dimensions as its input. Before feeding the image to the network, it undergoes preprocessing, which typically involves resizing the image to a predefined size and normalizing its pixel values to bring them within a specific range (e.g., [0, 1]).

Network Architecture:

The TinyYOLOv3 architecture consists of a series of convolutional layers, followed by a few fully connected layers. Unlike the original YOLOv3, TinyYOLOv3 has a reduced number of layers, resulting in fewer computations and lower memory requirements. While the model has fewer layers, it still maintains competitive accuracy for certain object detection tasks, including person detection.

Feature Extraction:

The convolutional layers in TinyYOLOv3 are responsible for feature extraction from the input image. These layers apply convolutional filters to the image, progressively learning to detect low-level features such as edges and textures, and then higher-level features representing more complex patterns and shapes.[7]

Detection:

TinyYOLOv3 performs object detection through the detection head, which predicts bounding boxes and associated class probabilities for the objects present in the image. The detection head is usually composed of multiple layers that process the features extracted by the earlier layers.

Output Prediction:

The output of TinyYOLOv3 is a grid of predictions that cover the entire input image. Each cell in the grid predicts a fixed number of bounding boxes and their corresponding confidence scores and class probabilities. The confidence score represents the model's confidence in the presence of an object within the bounding box, while the class probabilities represent the likelihood of the object belonging to a specific class (e.g., person, car, etc.).

Non-maximum Suppression (NMS):

To eliminate duplicate and overlapping detections, a post-processing step called non-maximum suppression (NMS) is applied. NMS filters out redundant bounding boxes based on their confidence scores and their overlap with other boxes. This ensures that each object is only detected once and retains the most accurate bounding box.[8]

Thresholding:

Finally, a threshold is applied to the confidence scores to filter out low-confidence detections. Objects with confidence scores below the threshold are discarded, resulting in a more confident and accurate set of detections.[9,10]

By following this process, TinyYOLOv3 efficiently and accurately detects objects, including persons, in real-time from input images and videos, making it a valuable choice for various applications where low-latency object detection is crucial.

IV. Testing and Results

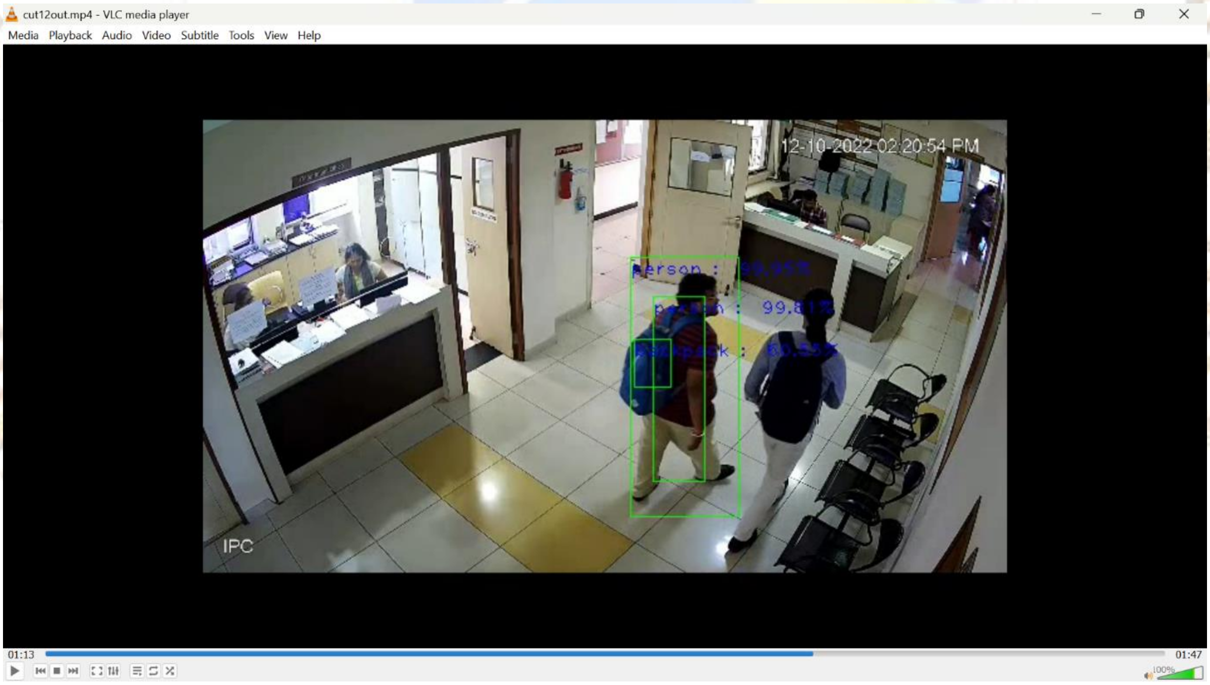

Fig 1.1 showing person being detected

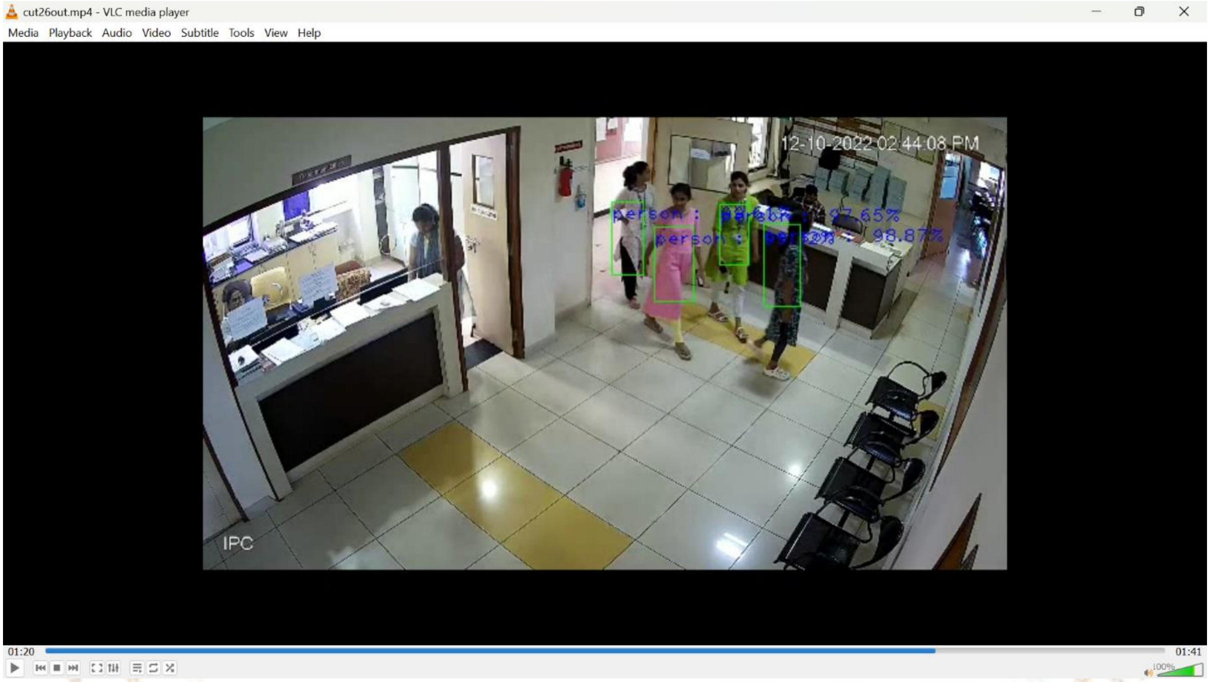

Fig 1.2 showing person detection when multiple persons are present in the frame.

V. DISCUSSION

TinyYOLOv3, a real-time object detection algorithm plays its role efficiently in performing object detection tasks, mainly person detection in this study. It is a lightweight version designed to work on resource constrained devices like smartphones and low power hardware. TinyYOLOv3 works well in predicting bounding boxes surrounding the object of interest. The model maintained faster inference speed while balancing accuracy. There were detection issues noticed when a person was present behind transparent material like glass which may have occurred due to change in refractive index. The model worked decent in detecting persons with darker skin tone which can be a plus point in real time detection. Even when larger group of people are present at a time in a frame, the model proved to be efficient in detection. It is also noteworthy to mention that tinyYOLOv3 detects persons irrespective of their position with respect to the camera.

TinyYOLOv3 has been known for its ability to perform real-time object (person) detection enabling high inference speeds as noted by Martinez-Alpiste et.al[11]. TinyYOLOv3's compact size allows it to operate efficiently with limited memory resources making it well-suited for deployment on edge devices where memory constraints are prevalent. It strikes a balance between model size and performance, making it an effective choice for scenarios where accurate detections are essential, but computational resources are limited. Being a one-stage object detection model, TinyYOLOv3 processes the entire input image in a single pass, eliminating the need for time-consuming region proposals and post-processing steps. This contributes to its real-time capabilities and overall efficiency, same as observed in the study conducted by Pranav Adarsh et.al.[12] The same study[12] also noticed that reduction in the depth of neural networks also reduced the accuracy rate compared to other YOLO versions which were slower. TinyYOLOv3 does possess other limitations like limited detection precision in crowded places, occlusion scenarios, limited feature learning capacity. It struggles with adapting to new domains and detecting smaller objects.

There are various approaches for identifying and localising objects, however they all trade off performance speed and accuracy. Due to the broad breadth of this research, object detection applications have quickly gained popularity. However, there is still more to be learned in this field. As research advances, more effective and accurate iterations of the YOLO series or other object detection algorithms have been created. When discussing object detection solutions, it's always a good idea to take the most recent research into account. The versions of YOLO have been upgraded to YOLOv4, YOLOv5, YOLOv6, YOLOv7, yoloV8.[13] Rather than declaring one algorithm to be superior to others one can always choose the approach that best fits the situation.

VI. CONCLUSION

In conclusion, TinyYOLOv3 offers numerous advantages, including real-time processing, efficient memory usage, and competent accuracy, making it a valuable option for person detection in resource-constrained environments. However, its limitations should be considered, particularly regarding precision, small object detection, and handling challenging occlusions. By understanding its strengths and weaknesses, developers can make informed decisions when choosing TinyYOLOv3 for person detection applications. In practical applications, TinyYOLOv3 demonstrates its value in various domains, including surveillance, robotics, and safety-critical systems. Its ability to perform real-time person detection, even on devices with limited computational power, empowers these systems with timely and effective responses.

As the field of computer vision continues to evolve, TinyYOLOv3 remains an important tool in the toolbox of object detection models, especially for those requiring efficient inference and real-time processing. By recognizing its strengths and limitations, and by adapting it to specific use cases, TinyYOLOv3 can play a pivotal role in enhancing safety, security, and interaction in our ever-evolving technological landscape.

VII. REFERENCES

1. YOLO realtime object detection algorithm. (2023).
2. Thangaraj, R. *et al.* Deep Learning based Real-Time Face Detection and Gender Classification using OpenCV and Inception v3. in *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)* 1–5 (IEEE, 2021). doi:10.1109/ICAECA52838.2021.9675635.
3. Ullah, R. *et al.* A Real-Time Framework for Human Face Detection and Recognition in CCTV Images. *Math Probl Eng* **2022**, (2022).
4. Redmon, J. & Farhadi, A. *YOLOv3: An Incremental Improvement*. https://pjreddie.com/yolo/.
5. Diwan, T., Anirudh, G. & Tembhurne, J. V. Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimed Tools Appl* **82**, 9243–9275 (2023).
6. Wilson Dawn, C Manusankar & P H, P. Analytical Study on Object Detection using Yolo Algorithm. *Int J Innov Sci Res Technol* **7**, 1 (2022).
7. Gong, X., Ma, L. & Ouyang, H. An improved method of Tiny YOLOV3. in *IOP Conference Series: Earth and Environmental Science* vol. 440 (Institute of Physics Publishing, 2020).
8. yolo object detector in Pytorch. (2023).
9. Gai, W., Liu, Y., Zhang, J. & Jing, G. An improved Tiny YOLOv3 for real-time object detection. *Systems Science and Control Engineering* **9**, 314–321 (2021).
10. Bandukwala, D., Momin, M., Khan, A., Khan, A. & Islam, Dr. L. Object Detection using YOLO. *Int J Res Appl Sci Eng Technol* **10**, 823–829 (2022).
11. Martinez-Alpiste, I., Golcarenarenji, G., Wang, Q. & Alcaraz-Calero, J. M. A dynamic discarding technique to increase speed and preserve accuracy for YOLOv3. *Neural Comput Appl* **33**, 9961–9973 (2021).
12. Adarsh, P., Rathi, P. & Kumar, M. YOLO v3-Tiny: Object Detection and Recognition using one stage improved model. in *2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020* 687–694 (Institute of Electrical and Electronics Engineers Inc., 2020). doi:10.1109/ICACCS48705.2020.9074315.
13. Joseph Nelson. Guide to YOLO models. (2021).

Figure legend:

Fig 1.1 - showing person being detected

Fig 1.2 - showing person detection when multiple persons are present in the frame.