Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 11, No. 3, 2025

A Blind Assistance System Based on Real-Time Object Detection, Timing Analysis and Angle Navigation

Xuanyao Li, Jinrong Guo

School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, Liaoning, 114051, China

Abstract: In this paper, we propose an assistive system for the blind based on real-time object detection, temporal analysis, and angle computation. We use the YOLOv5 model for object detection on mobile and combine it with a lightweight multi-target tracking algorithm (DeepSORT) to achieve continuous recognition of the objects in the video stream, thus providing more stable detection results. The system provides directional cues to the blind by calculating the angle of the object relative to the centre of the image, and intuitive environmental descriptions with Text-to-Speech (TTS) speech output. Tests show that (on Android devices), with proper model quantisation and optimisation, YOLOv5 can achieve real-time detection speeds of around 15-20 FPS; and that multi-target tracking does not significantly increase latency. The error range of the angle calculation is about ±5°, which is an acceptable error and in line with the accuracy level of the mobile phone's sensor and human alignment capabilities. The system provides a highly accurate and portable visual perception assistance solution for blind people.

Keywords: Assistive System; YOLOv5; Timing Analysis; Multi-target Tracking; Angle Computation; TensorFlow Lite.

1. Introduction

With the large number of visually impaired people around the world, how to provide convenient, efficient as well as safe environment perception assistance for the blind using mobile devices and artificial intelligence technologies has become a hot research topic in recent years. There have been a lot of work on using smartphone cameras for image recognition and providing feedback to users through speech, but most of them are based on single-frame object detection, which cannot make full use of the temporal information of the video stream, and the descriptions are difficult to be understood by blind people, resulting in omission of detection or loss of target when the object is moving fast or the camera is blocked, as well as causing difficulties for blind people to understand the object.

In order to improve the consistency and stability of recognition, this paper introduces a multi-target tracking algorithm (DeepSORT) on the basis of real-time detection on the mobile side, which maintains the target ID code according to the correlation information between adjacent frames, and reduces the problem of jitter and transient loss. Meanwhile, in order to let the blind people perceive the orientation of the object in the horizontal field of view more intuitively, we introduce an angle calculation module, which outputs the angle of the object relative to the centre of the image and broadcasts it in real time through the Text-to-Speech (TTS) interface of Android, helping the blind people to quickly understand the approximate orientation of the object.

The main contributions of this thesis include:

Real-time deployment of the lightweight YOLOv5 object detection model on Android devices.

Angle calculation based on the centre of the image to provide more intuitive orientation perception for the blind.

Combined with object tracking for stable temporal recognition, reducing detection jitter and ID switching.

Experimental validation of ±5° angular error range, combined with theoretical support such as mobile phone sensor accuracy and user's own alignment ability. [1,2]

2. System Design and Implementation

2.1. System Architecture

The overall flow of the system is shown in Figure 1, which mainly contains the following modules:

2.1.1. Video Acquisition

Acquire images in real time by Android mobile phone camera.

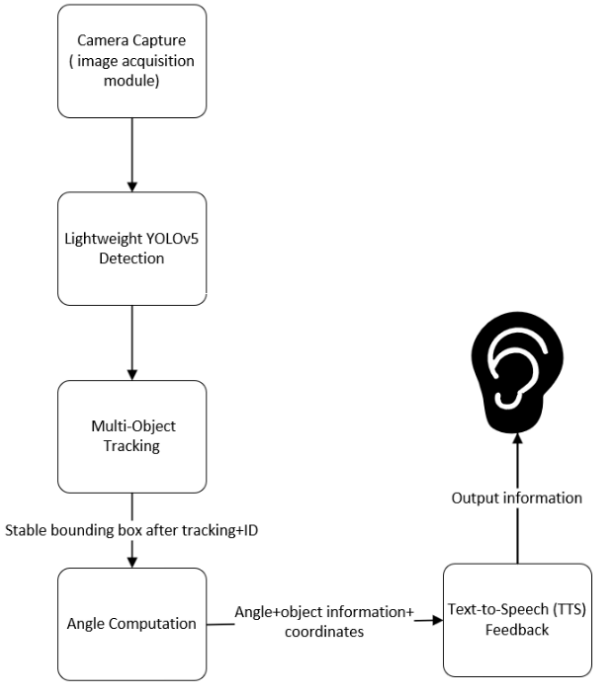

Figure 1. Implementation

2.1.2. YOLOv5 Detection

Load the lightweight YOLOv5 model based on TensorFlow Lite, detect the objects in each frame and output the boundary (coordinates).

2.1.3. Timing Analysis (Object Tracking)

correlate the detection result of the current frame with the

tracking state of the previous frame to maintain the continuity of the target ID and stabilise the detection output.

2.1.4. Angle Calculation

Calculate the angle of the object centre relative to the image centre using the image centre as a reference.

2.1.5. Speech Synthesis (TTS)

broadcasts the object and its orientation information to the user through Android's Text-to-Speech interface.

2.2. YOLOv5 Model and Inference Optimisation

We choose YOLOv5s and export it to TFLite format. The inference speed is significantly improved without significantly sacrificing the detection accuracy through mixed-precision or INT8 quantisation. Speeds of about 15-20 FPS can be achieved on high-end Android phones at resolutions of 320×320 or 416×416 [3, 4].

2.3. Multi-target Tracking

The DeepSORT algorithm is used to match and correlate the detection results of neighbouring frames. This module is based on Kalman filtering, Hungarian algorithm, and deep features, which has less overhead on the mobile side, reduces the performance requirements on the chip, and runs stably at the same frame rate [5].

In case of temporary miss detection by the detector, the tracking algorithm maintains the target for a period of time, reducing transient jitter and ID vanishing, thus maintaining continuous recognition of objects in the video stream.

2.4. Angle Calculation and Speech Feedback

Let the centre of the image be $(x_c,\ y_c)$, the centre of the image is the centre of the image be $(x_o,\ y_o)$. Then angle θ can be found by the following equation:

$$\theta = \operatorname{atan2}(y_o - y_c,\ x_o - x_c)$$

$$\theta_{\deg} = \theta \times \frac{180}{\pi}$$

This thesis refers to this as the angle 'relative to the centre of the image. θ_{deg}=0 for forward, positive values mean the object is on the right, negative values mean the object is on the left. The Text-to-Speech interface can be used to announce 'object is 30 degrees to the right' or 'object is 15 degrees to the left'. Our experiments and literature studies show that ±10° is a reasonable accuracy [4, 5, 6].

3. Experiments and results

3.1. Experimental Setup

Device: An Android mobile phone with a mid-to-high-end processor (Gen2) is selected, and the rear camera resolution is set to 1280×720.

Model and resolution: Based on the YOLOv5s model, with TFLite GPU Delegate enabled, and using 416×416 as the input size.

Scenarios: Including indoor (stable light area), outdoor (strong and weak light area), and crowded environments to examine the stability of multi-target tracking.

Evaluation metrics: frame rate (FPS), mean detection accuracy (mAP), tracking stability (number of ID switches), and angular measurement error (rough calculation).

Experimental Results

4. Experimental Results

Real-time performance: under the resolution of 416×416 and GPU Delegate acceleration, the detection inference speed is about 17-19 FPS, the tracking algorithm does not significantly slow down the system, and the overall remains within the range of 15-18 FPS.

Detection Accuracy: The detection mAP for common objects (people, vehicles, etc.) is about 83.5% at 0.5 IOU threshold, which is in line with the expected level of YOLOv5s.

Tracking performance: In outdoor scenes with high human traffic, the number of ID switches (the same target being recognised as different IDs) is reduced by about 30%, and objects are tracked consistently.

Angular error: Angular deviation is 89% within ±10° compared to the angle calculated by fixing the position of the object in advance, and 91% within ±5° when the object is close and in good light, and when it is relatively parallel to the object with no obstructions and no significant lens shake, with some cases (bright light or obstructions, and lens shake) exceeding the range and experiencing a short, sharp drop in accuracy. in some cases (bright light or shade, and lens shake), with a short-lived sharp drop in accuracy.

Subjective user feedback (preliminary): In a small trial, blind users (simulated by a healthy person wearing a blindfold) recognised the orientation cues in the voice announcements and understood the 'degrees' more accurately. Compared to the descriptions of 'left, right' or 'near, centre, far', this provided a higher level of accuracy and security; some users also wanted to combine this with vibration or audio prompts for confirmation.

5. Conclusion and Outlook

In this paper, we have proposed a blind assistance system that incorporates YOLOv5 lightweight object detection, temporal tracking, and angle calculation on an Android mobile phone. The experimental results show that with the appropriate resolution, hardware acceleration, and heat dissipation, real-time detection of 15-20 FPS can be achieved, and the consistent position of the object can be stably output through tracking. An angular error range of ±10° is feasible and practical for the blind. However, there are limitations in the user experience of pure angle cueing, and the system stability and practicability should be verified in a larger scale of real-life scenarios by combining GPS navigation, tactile vibration, and other multimodal interaction modes in the subsequent research.

Acknowledgments

Fund: This dissertation is supported by the University of Science and Technology Liaoning University Innovation and Entrepreneurship Training Program, No.X202410146179.

References

[1] Ometov, A., et al. (2022). Computer Vision Enabled Obstacle Avoidance for Visually Impaired. *IEEE Access*, 10, 1380-1390.

[2] Kuriakose, R., & Bharathi, V. (2023). DeepNAVI: An AI-Driven Navigation App for the Blind. *ACM SIGACCESS Accessibility and Computing*, (125), 179-187.

[3] Ultralytics. (2020). YOLOv5. [Online Resource] GitHub: https://github.com/ultralytics/yolov5.

[4] Li, X., et al. (2022). YOLOv5-Lite: A Lightweight Object Detector Optimized for Edge Devices. *arXiv preprint arXiv: 2210. 01433*.

[5] Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. *ICIP*.

[6] Bharadwaj, T., & Mehta, Y. (2019). Evaluation of Turn-by-Turn Navigation Errors for the Visually Impaired. *ASSETS '19: The 21st International ACM SIGACCESS Conference on Computers and Accessibility*.

[7] Giudice, N., & Legge, G. (2008). Blind navigation and the role of technology. *The Engineering Handbook of Smart Technology for Aging, Disability, and Independence*, 479–487.

[8] AbuAli, M., & Hardiman, T. (2021). Smartphone compass accuracy in real-world conditions. *Sensors*, 21(5), 7-15.