An Efficient and Reliable Real-Time Obstacle Detection System Using YOLO11 and SGBM

Tongle Yao, Xuedong Pan, Siyi Gao, Yufei Huang
Instructor: Prof. Jeongkyu Lee, Khoury San Jose

Northeastern University
Khoury College of Computer Sciences

Introduction

The development of descriptive vision assistance for visually impaired individuals is a key focus in computer vision and assistive technology. Current solutions, like Apple's door detection feature [1], rely heavily on continuous cloud connectivity or expensive hardware such as LiDAR, making them less accessible.

To address this, we designed a ***pure vision based***, ***lightweight***, and ***real-time*** obstacle detection system using ***YOLO11*** and ***Stereo Disparity Using Semi-Global Block Matching (SGBM)***. Our solution is ***open-source***, deployable on personal devices with cameras, and focuses on reducing computational complexity while maintaining high accuracy.

Problem Statements

- **Limited Dataset Relevance:** Existing object detection models are trained on datasets like COCO, which lack relevance for visually impaired users.
- **Imbalance in Public Datasets:** Many public datasets contain excessive irrelevant or insufficiently diverse data.
- **Lack of Depth Information:** Most models only classify objects but fail to estimate their distance from the user.
- **Performance Constraints:** Real-time applications are impractical on personal devices due to high computational requirements.

Methodologies

Dataset Preparation: We utilized the ***Objects365*** and ***Cityscapes*** datasets to achieve rich category coverage and diverse image quality. We selected 42 of the most common obstacles encountered by visually impaired individuals, as shown in Fig. 1. To refine the dataset, we performed ***data cleaning***, applied ***data augmentation*** techniques, and used ***weighted sampling*** to address class imbalances, with these steps outlined in Fig. 2. The result was our final optimized and balanced dataset, shown in Fig. 1.

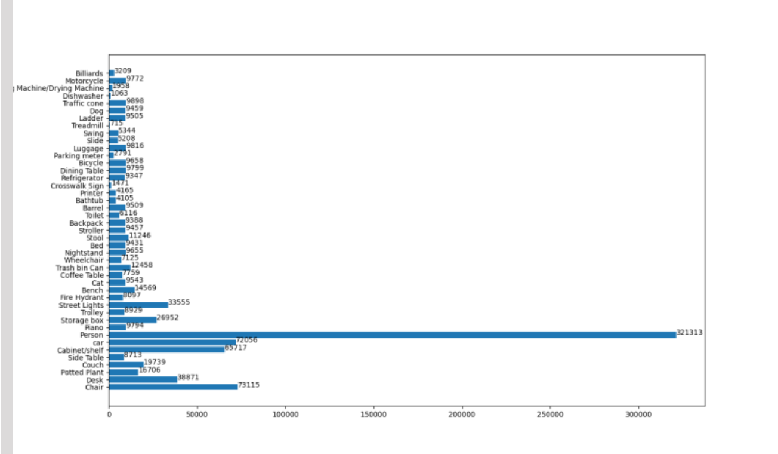

Fig. 1. Balanced Data Distribution of 42 Selected Classes

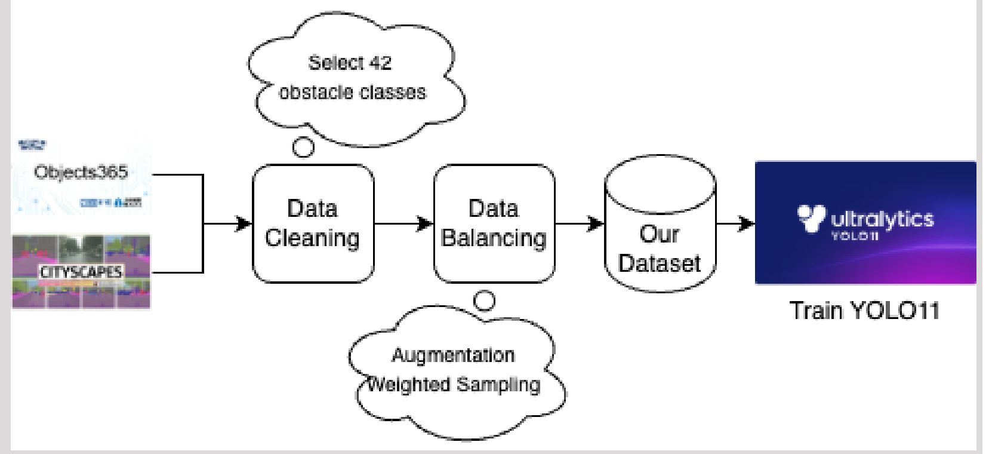

Fig. 2. Data Pre-processing Workflow

Object Detection: We selected YOLO11 for its enhanced accuracy, lower computational demands, and edge-device compatibility, making it ideal for real-time object detection in our application. YOLO11's architecture (Fig. 3) integrates a ***Backbone*** for efficient feature extraction, a ***Neck*** for feature fusion and multi-scale optimization, and a ***Head*** with auto-anchor capabilities to improve detection across various object sizes.

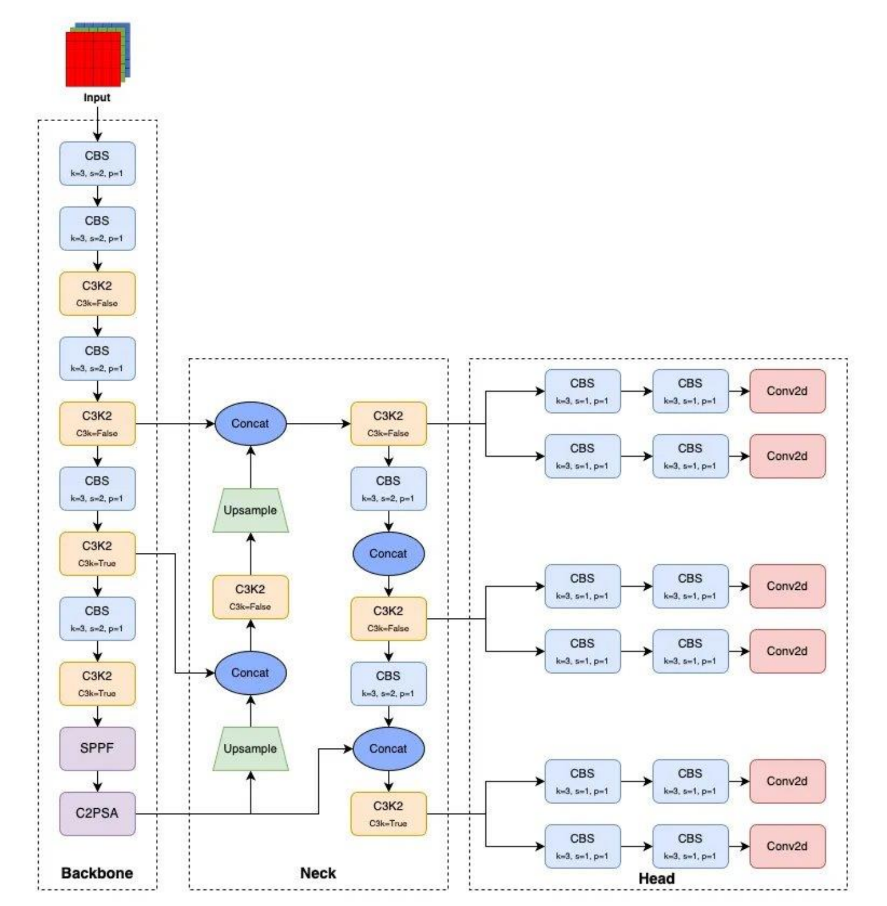

Fig. 3. YOLO11 Architecture [2]

We trained both the ***YOLO11-nano*** and ***YOLO11-s*** models on separate sub-datasets, employing parameter optimization strategies to ensure stability and generalization. Additionally, ***mixed precision training*** and ***real-time monitoring*** were used to evaluate performance and prevent overfitting.

Distance Estimation: We chose the ***SGBM*** algorithm for stereo vision depth estimation because of its balance of accuracy, real-time performance, and low computational requirements. It calculates disparity from stereo images, refines it with semi-global optimization, and converts the results into 3D coordinates. Optimizations like ***multi-point depth sampling***, ***region-specific computation***, and ***down sampling*** improved speed tenfold, ensuring real-time performance.

Text-to-speech (TTS): We implemented TTS using the ***Pyttsx3*** library to provide real-time audio notifications by prioritizing the closest object to the user, alerting visually impaired users about potential hazards.

Fig. 4 summarizes and illustrates the overall workflow of our program.

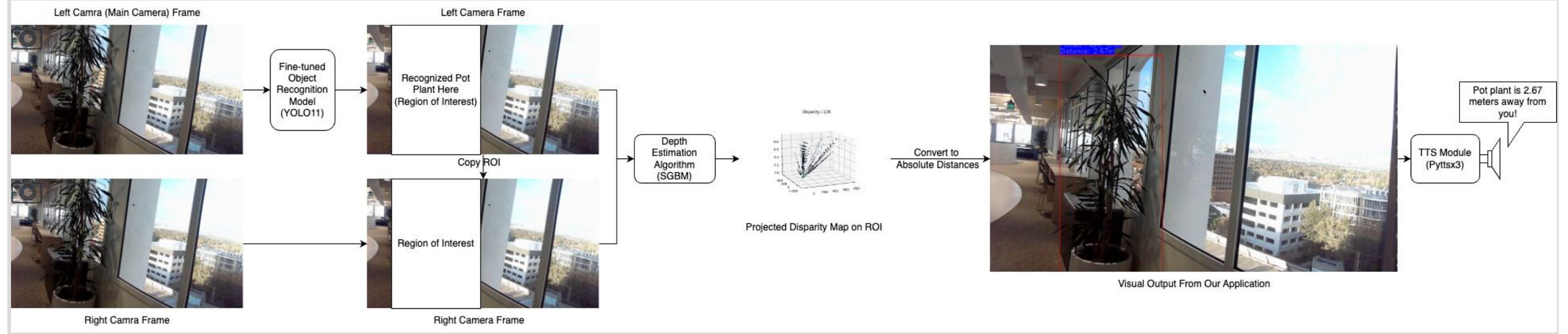
Left Camra (Main Camera) Frame
Fine-tuned Object Recognition Model (YOLO11)
Left Camera Frame
Recognized Pot Plant Here (Region of Interest)
Copy ROI
Region of Interest
Right Camra Frame
Right Camera Frame
Depth Estimation Algorithm (SGBM)
Projected Disparity Map on ROI
Convert to Absolute Distances
Visual Output From Our Application
TTS Module (Pyttsx3)
Pot plant is 2.67 meters away from you!

Fig. 4. Overall Program Workflow

Results

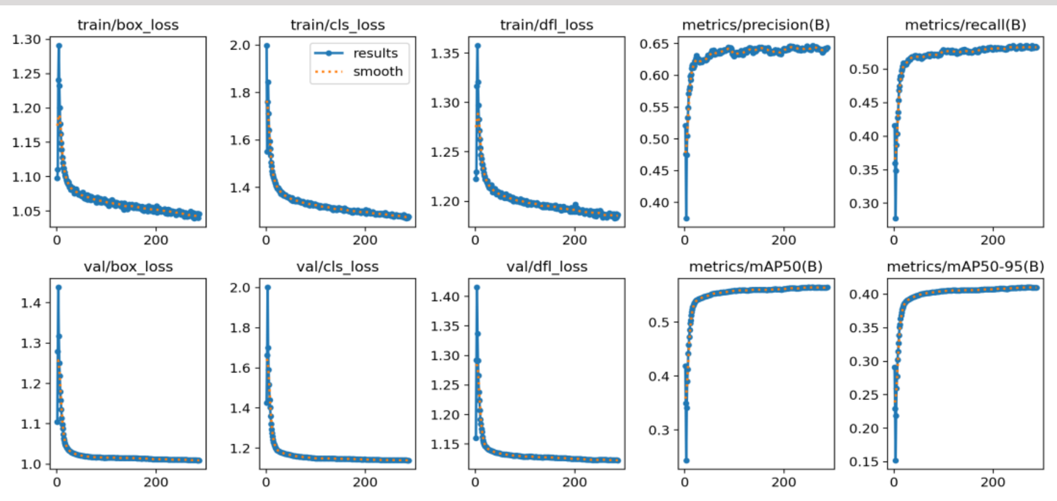

Fig. 5. Model Evaluation Graphs

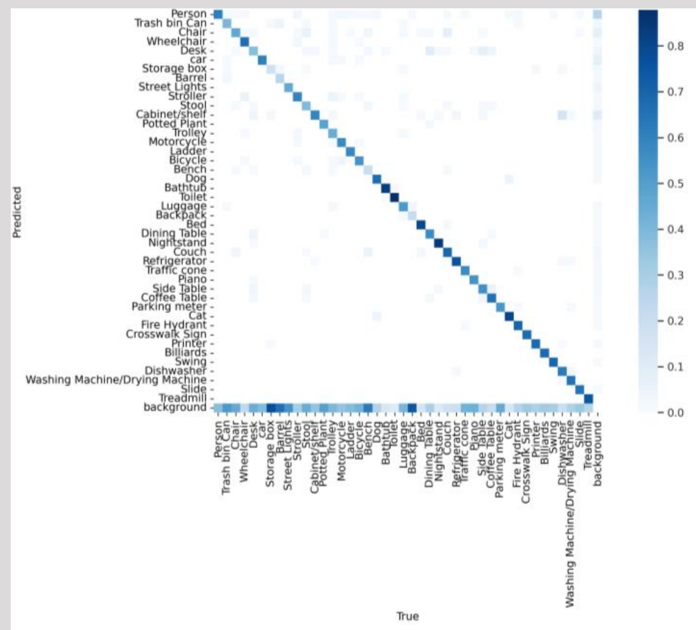

Fig. 6. Confusion Matrix of All Classes

Object Detection Evaluation: The model showed consistent improvement in accuracy, with key metrics such as mAP and Recall significantly increasing throughout the training epochs, ultimately reaching **56% mAP@50** and **41% mAP@50-95**.

The confusion matrix revealed minor misclassifications, particularly between visually similar objects (e.g., "Chair" vs. "Desk" in Fig. 6).

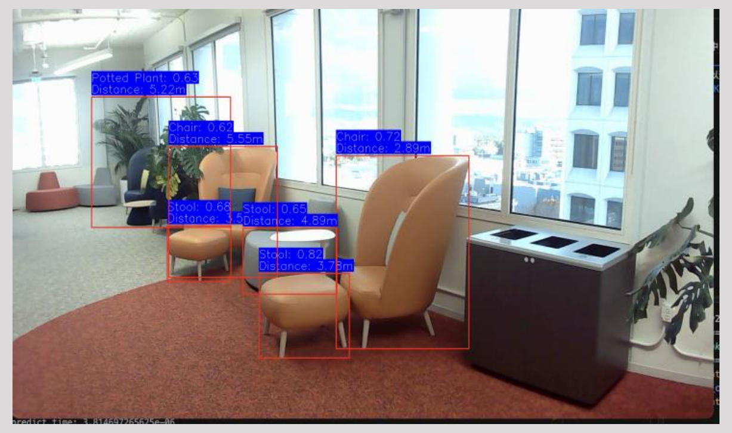

Fig. 7. Example Visual Outputs from Our Program

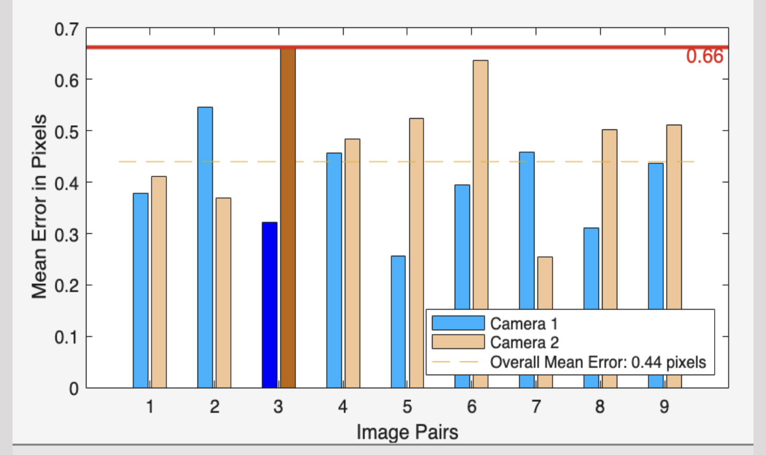

Fig. 8. Example Visual Outputs from Our Program

Distance Estimation Performance: The SGBM algorithm achieved reliable relative distance measurements in Fig. 7, maintaining a ***real-time*** frame processing rate of around ***20 fps*** in our demonstration [3]. Optimization challenges led to higher-than-expected Mean Square Pixel (MSP) error, likely due to camera calibration limitations shown in Fig. 8

Conclusion and Future Works

Our system successfully combines YOLO11 and SGBM to deliver efficient real-time obstacle detection and distance estimation. Future work includes:

- Deploying the system on low-cost devices like Raspberry Pi with Intel Neural Sticks.
- Enhancing distance accuracy through improved camera calibration.
- Exploring semantic segmentation for irregularly shaped objects.

References

[1] Apple Inc., "Accessibility," 2024, Apple. [Online]. Available: https://www.apple.com/accessibility/

[2] N. Jegham, C. Y. Koh, M. Abdelatti, and A. Hendawi, "Evaluating the Evolution of YOLO (You Only Look Once) Models: A Comprehensive Benchmark Study of YOLO11 and Its Predecessors," 2024, arXiv:2411.00201. [Online]. Available: https://arxiv.org/abs/2411.00201

[3] Uni 404, "CS5330 Group5 Final Proj Demo," YouTube, Nov. 23, 2024. https://www.youtube.com/watch?v=9njp5Nq8DAc

CS 5330 Pattern Recognition and Computer Vision in Fall 2024