Visually Empowered: Real-Time Object Recognition and Auditory Description System

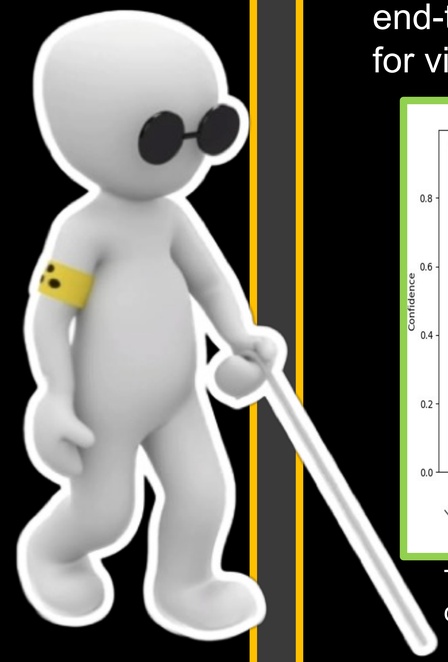
Kundhana H. Paruchuru, Department of Computer Science, Indiana University, Bloomington

Motivation

- Technologies used
The seamless integration of computer vision, NLP, and text to speech conversion.

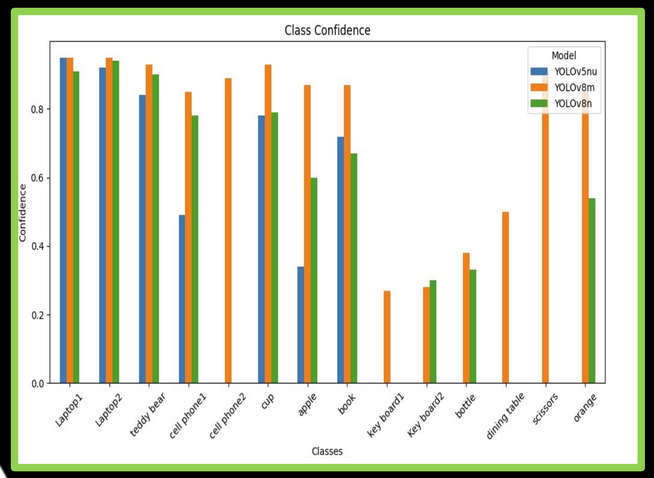
- Multilingual Capabilities
Breaking language barriers and making information accessible to a global audience.

- Impact and Accessibility
Benefit people with visual impairments and those who prefer audio content, enhancing their access to information and experiences



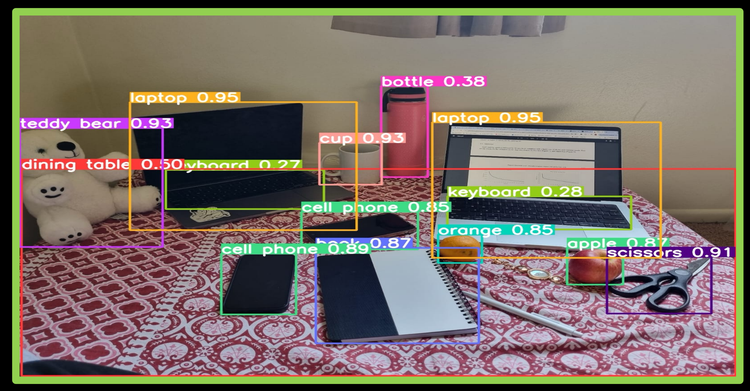
Summary

- Our research focuses on developing an end-to-end system to enhance accessibility for visually impaired individuals.

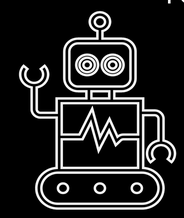


The bar graph displays the confidence levels for different classes predicted by various models.

Results



YOLOV8m best performing model



There are 2 laptops, 2 cell phones, 2 keyboards and teddy bear, cup, scissors, book, apple, orange, dining table, and bottle.

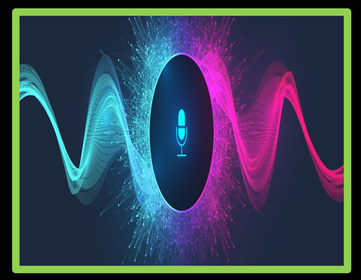
Hay 2 computadoras portátiles, 2 teléfonos celulares, 2 teclados y un osito de peluche, una taza, tijeras, un libro, una manzana, una naranja, una mesa de comedor y una botella.

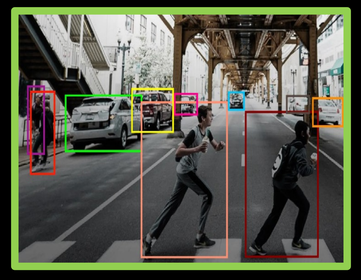
Есть 2 ноутбука, 2 мобильных телефона, 2 клавиатуры и плюшевый мишка, чашка, ножницы, книга, яблоко, апельсин, обеденный стол и бутылка.

Text generated for this image in 3 different languages

Methods

- Different versions of pretrained YOLO are chosen for efficient real-time object detection





Voice Integration and Object detection using COCO dataset

- Utilized Google Translate API within an NLP pipeline to convert into descriptive text across 55 languages.

Data Collection

- Selected COCO dataset due to its comprehensive annotations and diverse object classes
- Comprehensive dataset with 80 distinct object classes.
- Classes cover a diverse range of everyday objects, animals, vehicles, and more.

COCO dataset

Future Work

- Integration with wearable and IOT devices
- Intelligent interaction and feedback

References

- Susitra, K., Dineshsakthi, M., & Krishna, V. S. (2023, December). Various Technologies on Object Detecton and Natural Language Processing for Visually Impaired People. In *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)* (pp. 1-4). IEEE.