# SafeWalk

## *for the Visually Impaired*

Empower visually impaired individuals with the safety, independence, and confidence to navigate the world.

# Our Team: Innovators in Accessibility

**Kris Junghee Lee**

Data Collection, Yolo Fine tuning, Alert Rule Setting , App Test & Development

**Helen Hu**

Datasets Collection & Cleaning, Model Evaluation

**Ke Zhang**

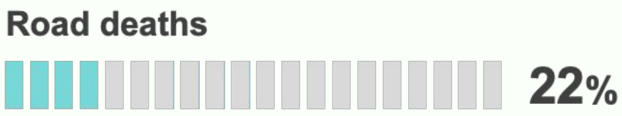
Data pipeline, VLM, App Test and Webmaster

**Ula Zhu**

Project Manager, UI/UX design

# Problem Statement

- **Over 36 million** people globally live with severe visual impairments, which is expected double in the next 30 years
- **More than 270,000** pedestrians are killed on roads each year




TRANSPORTATION

Disabled People Are Dying in America's Crosswalks. We Need to Protect Them.

The data on traffic fatalities and injuries doesn't account for their needs or even count them. Better data would enable better solutions.

OPINION | March 27, 2024 • Claudia Folska

Diego Servo/Shutterstock

Latest News

FINANCE
Under Trump, Medicaid Faces a 'Reset Moment'
Jan. 13, 2025

WORKFORCE
Minimum Wages Are Rising in Nearly Half the States This Year
Jan. 13, 2025

POLITICS
Jimmy Carter's Redemptive Record on Race and Injustice
OPINION | Jan. 13, 2025

ARTIFICIAL INTELLIGENCE
Health-Care AI Requires a Lot of Expensive Humans
Jan. 13, 2025

NEWS IN NUMBERS
10
January 10, 2025

# Target Users

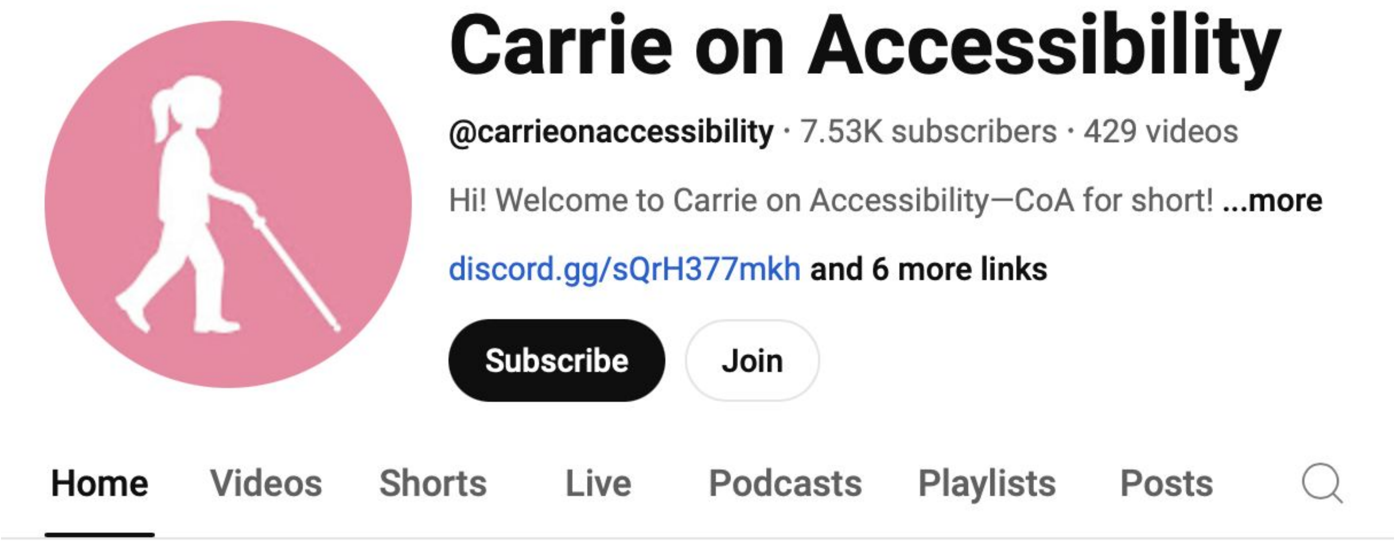
Visually impaired individuals living in urban areas with smartphone accessibility


# Market Size

2021
**$3.9B**

2030
**$13.2B**

# Interview with Domain Expert



# SafeWalk Key Features

### Obstacle Detection

Identify outdoor obstacles

Landscape mode

### Adaptive Warning

*"Be careful"* for close objects

*"In the front"* for further objects

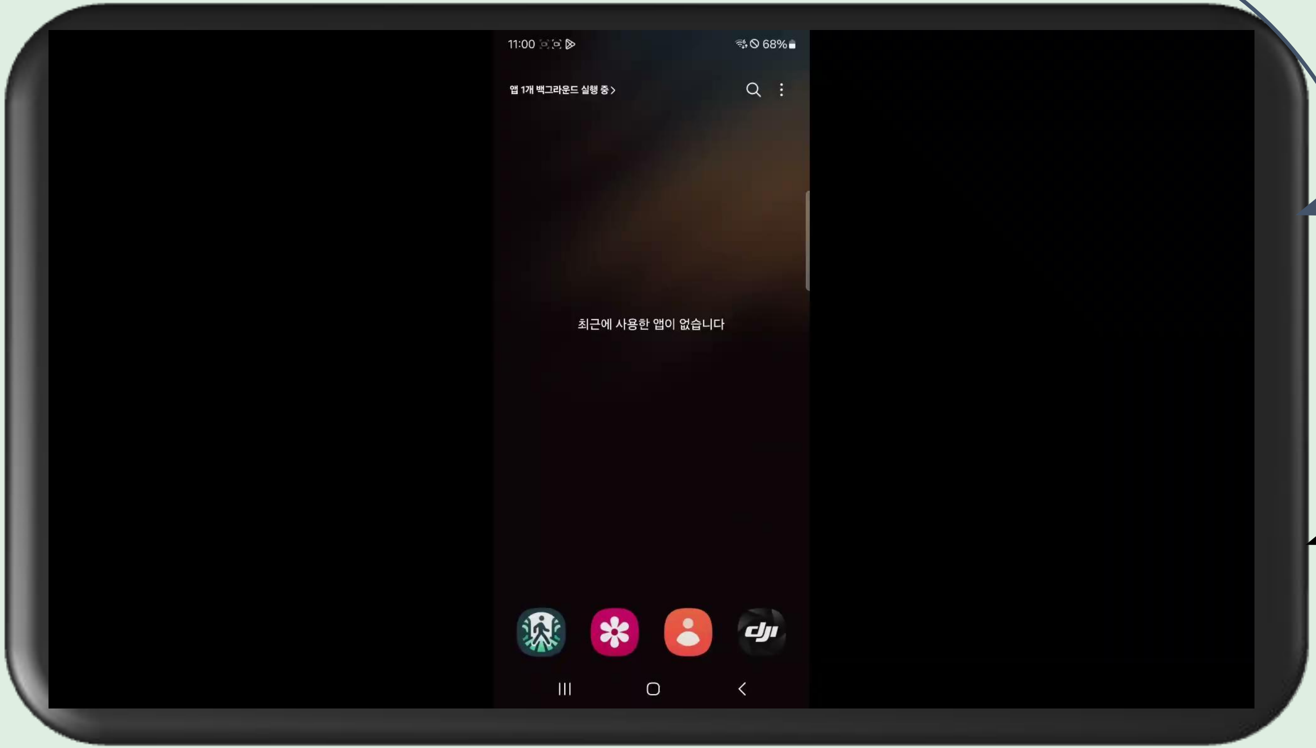
**Direction:** front, left, right

### Surrounding Description

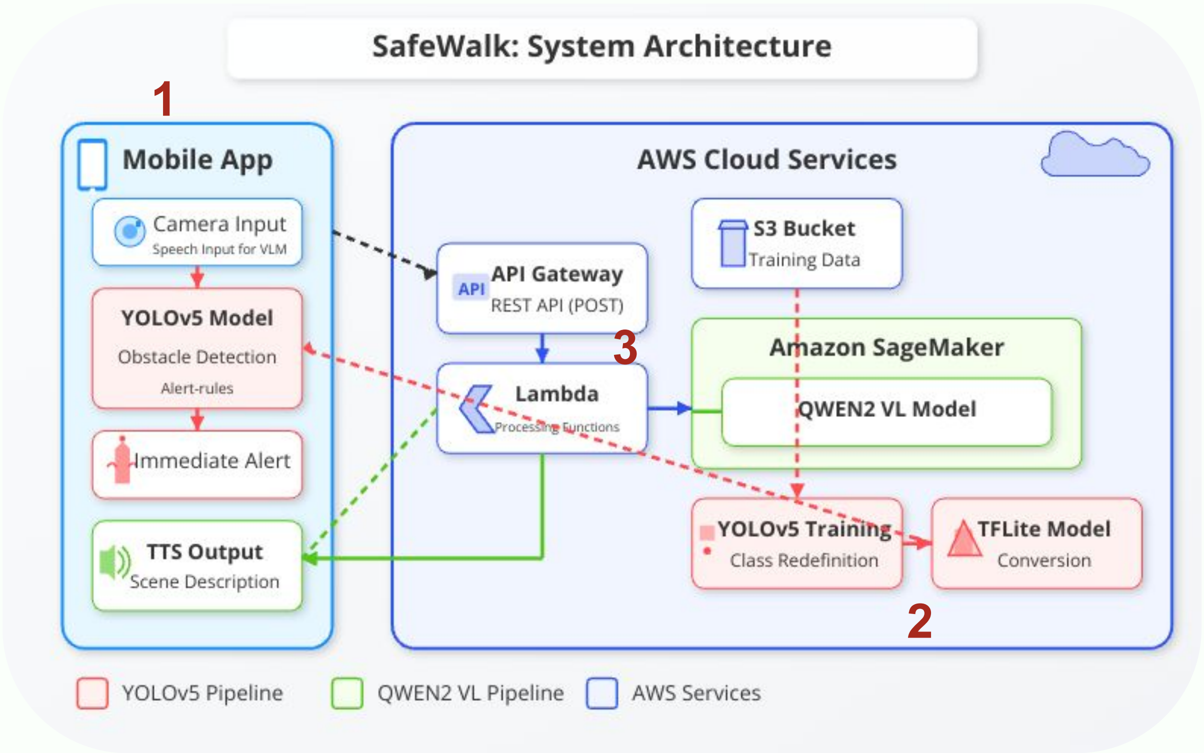
Helps users understand what's in front of them real-time.

### Voice Assistant

Answer questions



# System Architecture Design




## Mobile App Front-End

- Real-time Obstacle Detection & Alerts
- Scene Description via VTT and TTS Output
- Flutter was used to make front-end

## YOLOv5 Pipeline

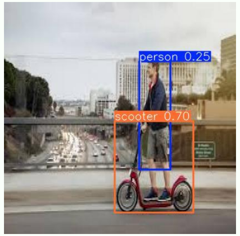
- Fine-tuned on Custom Dataset
- Converted to TFLite for Mobile

## Qwen2-VL Pipeline

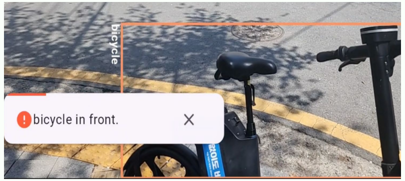
- Custom Prompt Engineering for Accessibility
- Lambda + API Gateway for Mobile Integration

# Overcoming Technical Challenges

*“scooter”*




*“Be Careful Bicycle in front”*




*“The image shows a city street scene during the day ...”*

## 1

### Identify Obstacles

- Train dataset
- Model selection
- Accuracy

## 2

### Alert Rules

- Usability
- Informative warning

## 3

### Scene Description

- VLM Model Evaluation

# Build Unique Datasets with Real-World Examples



## Data Classes

- **Common (14 classes)**

  Person, Bicycle, Car, Motorcycle, Bus, Truck, Traffic Light, Fire Hydrant, Stop Sign, Parking Meter, Bench, Bird, Cat, Dog

- **Additional (13 classes)**

  **Pole, Pothole, Tree, Scooter, Bollard, Traffic Cone, Bad Roads, Fence Barrier, Stairs, Trash Bin, Downstairs, Upstairs, Crosswalk**

## Challenges

**Required many dataset due to Catastrophic Forgetting of Pre-trained model**

**Missing labels from Public images**

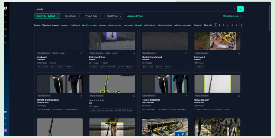
**Different images between Public Vs. Real-World**



## Solution

**COCO dataset**

roboflow

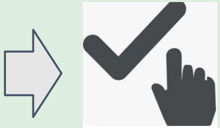


kaggle

**Auto Label Function**


Wrong-Auto Labeler Example

**Human Final check**



- **Self-taken & labeled real-world images**
- **Data augmentation**

Augmentation

**Final Dataset: 90,855**

# Selecting the Base Model to Reach Our Goal



## Performance-Oriented Model Comparison

| | Precision | Recall | mAP 0.5 | mAP 0.5-0.9 |
|---|---|---|---|---|
| Yolo v5s | **0.65** | **0.48** | **0.52** | **0.31** |
| Yolo v7s | 0.63 | 0.33 | 0.37 | 0.20 |
| Yolo v11s | **0.66** | **0.53** | **0.56** | **0.37** |

- mAP 0.5: Mean Average Precision calculated using a single Intersection over Union (IoU) threshold of 0.5.
- mAP 0.5–0.9: Mean Average Precision averaged over multiple IoU thresholds, ranging from 0.5 to 0.95 in increments of 0.05.

## Deployment-Focused Comparison

Well-established as a ***lightweight version, Robust compatibility*** with various deployment

Higher accuracy but ***slower inference***
Need ***Additional Effort*** for compatibility

**While Performance metrics show only minor differences, Yolo v5s shows substantial competitiveness in deployment**

☑

**Final Model Approach**

**Yolo v5s + Fine tuning**

# Optimizing performance with Fine-tuning

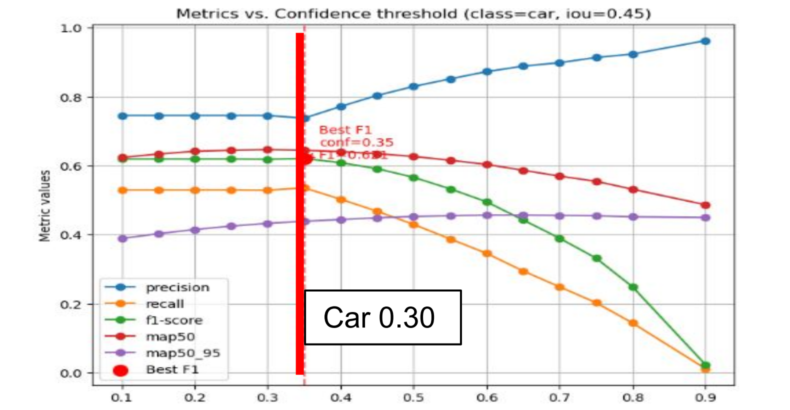


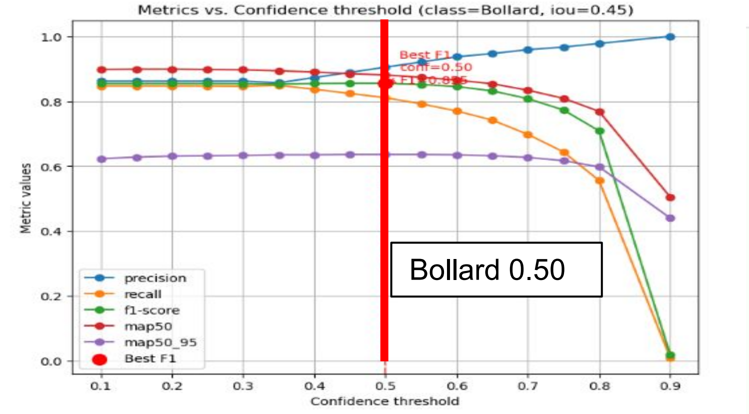
## Fine Tuning Approach

- Change optimizer (SGD-> Adam)
- Adjusted Learning Rate (0.01 -> 0.001)
- Increased Epochs (100)
- Class Weight Parameter
- Under sampling
- Oversampling by data augmentation
- Apply relative Threshold by class

- SGD: Stochastic Gradient Descent
- Adam: a method for stochastic optimization

## Example of threshold adjustment



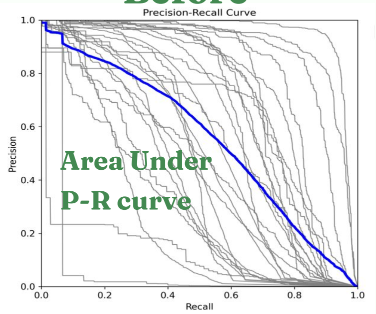
# Optimizing performance with Fine-tuning

## Final Evaluation

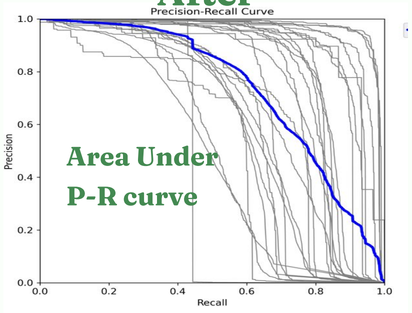
| | Precision | Recall | F1 Score | mAP 0.5 | mAP 0.5-0.95 |
|---|---|---|---|---|---|
| **Basemodel** | 0.65 | 0.48 | 0.54 | 0.52 | 0.31 |
| **Fine-Tuned model** | **0.82** | **0.67** | **0.73** | **0.76** | **0.54** |

## Precision Recall Curve

### Before



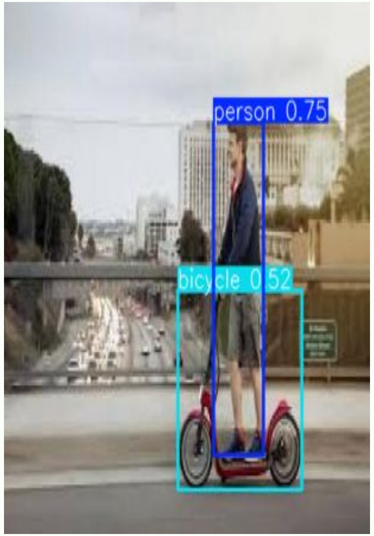

### After




iOU Threshold:0.5

## Object Detection Example

### Before




### After

# Building Adaptive Alert for better usability



## Rule Development using Bounding Box Parameter


*"Be Careful Bicycle in front"*

## Key Objective

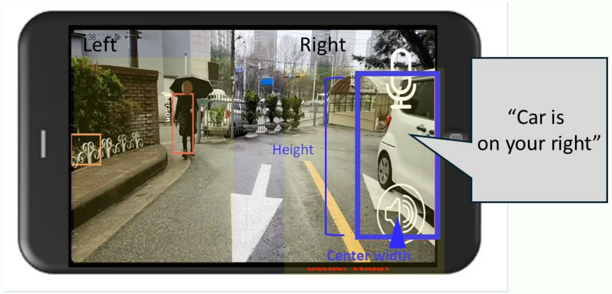
- Only alerts when it matters to ***reduce fatigue***
- Directional guidance for ***enhanced usability***
- Right-time warnings for ***safer decisions***

# Building Guided Alert for better usability

## Adaptive Alert

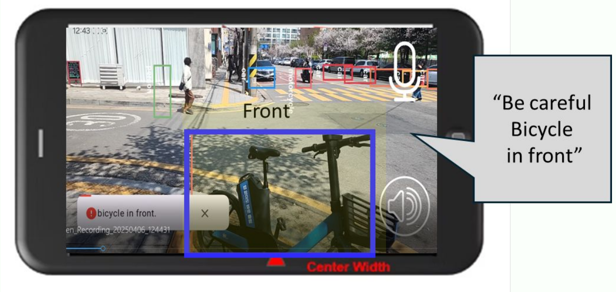
### Detect Mode (Directional Guide)

- Parameter : Center Width, Height, Width



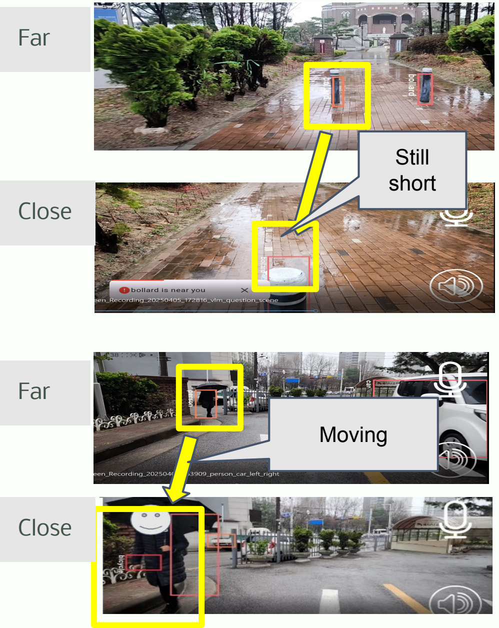

### Warning Mode (Be careful)

- Parameter: Center Width, Height, Width,Top, Bottom




## Challenges

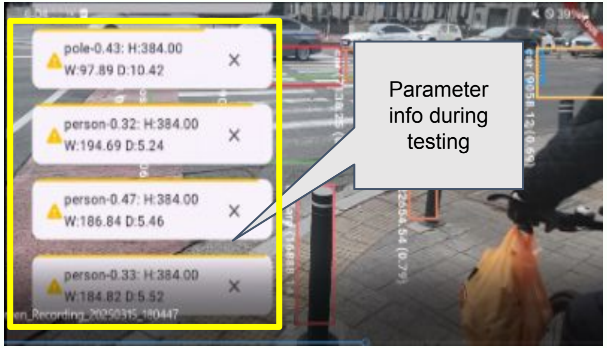
Uniform parameter conditions are not applicable,as each class differs in relative size –from short (bollard) to tall (tree), wide (car), and dynamic objects like moving persons.




## Approaches

Extensive testing and iterative adjustments with debugging test mode and as we classified test images for each class, configured parameter conditions

<Test Mode>



1 Identify Obstacles | 2 Alert Rules | 3 Scene Description

# Find the Model that Best Explains Scenes to Users

## VLM Evaluation- Example


**Q: Please describe what is in front of me**

**Paligemma's description:**

"The image is a video of a crosswalk with a person and a car in the middle of the road."

**Qwen2's description:**

“The image shows a city street scene during the day time. There are several people crossing the street at the pedestrian crossing marked with yellow line. The street is relatively wide and has a few cars parked along the side. There are trees with blooming flowers likely cherry blossoms..."

| Model | Navigation Safety (Score & Explanation) | Spatial Orientation (Score & Explanation) | Environmental Awareness (Score & Explanation) |
|---|---|---|---|
| Paligemma | 2.5/5<br>• Mentions person & car in road<br>• Vague on safety context | 1/5<br>• Only says “crosswalk”<br>• No layout or direction | 1/5<br>• No weather/time info<br>• No surroundings |
| Qwen2 | 3/5<br>• Notes pedestrian crossing<br>• Mentions people walking | 4/5<br>• Describes street & parked cars<br>• Mentions crosswalk & layout<br>• Good scene structure | 3.5/5<br>• Mentions trees & cherry blossoms<br>• Notes daytime setting<br>• Adds calm ambiance |

WIN!

1 Identify Obstacles
2 Alert Rules
3 Scene Description

# VLM Evaluation- Overall Result

| Model | Navigation Safety | Spatial Orientation | Environmental Awareness | Overall |
|---|---|---|---|---|
| Goal | Identify hazards, obstacles, and safe paths | Help users build a mental map | Context about surroundings for navigation decisions | — |
| Paligemma | 3.0 | 2.1 | 2.2 | 2.4 |
| Qwen2 | 3.9 | 3.8 | 3.9 (WIN!) | 3.9 |
| Gap | +0.9 | +1.7 | +1.7 | +1.4 |

WIN!

**Qwen 2**
**vs**
**Paligemma**

*Evaluated using 30 real-world urban scene*

**67%** better performance across all metrics

**4-6x** **more detailed descriptions**

(avg. 65-70 words vs. 11-12 words)

# Metrics on VLM captions evaluation

Compare text similarity between human descriptions and model generated descriptions

| | ROUGE | | | METEOR |
|---|---|---|---|---|
| | Rouge 1 | Rouge 2 | Rouge L | Meteor |
| **Qwen2** WIN! | 0.40 | 0.15 | 0.27 | 0.36 |
| **Paligemma** | 0.21 | 0.07 | 0.16 | 0.20 |

**Sample:** 30 images from 12 different cities in 7 countries

**ROUGE:** Share of overlapping words and phrases of different lengths between human descriptions and model descriptions

**METEOR:** Harmonic mean of unigram precision and recall, with recall weighted higher than precision

# Key Achievements

### Customized model

- New classes on outdoor objects with large training datasets
- High accuracy at relatively reasonable training costs

### Informative alerts

- Box-boundary based alerts
- Additional colors on the position of objects
- Combined with VLM for detailed descriptions

### Efficient deployment

- Yolo model deployed **on-device** to reduce latency
- VLM model via **API**

# Next Steps: Expand Vision Capability

### Expanding OS Compatibility

Finalize a version of the App for IOS system

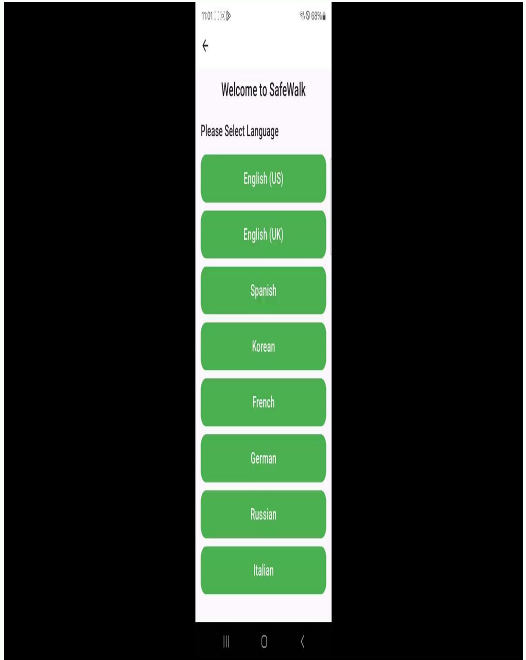
### Alert Rule with ML Classification Model

Use machine learning method to optimize alerts based on the object and screen size of users' devices

### Expanding Global Services

Customize object detection models for specific countries or cities

Enhance multilingual services (Quick Spanish demo)




### Wearable Device Integration

Connect with wearable cameras such as Go-Pro, Insta360 Go3, for hands-free options

Look for more cost-efficient solutions

# A World Equally Accessible to All

### Our Problem

Difficulty of visually impaired individuals when navigating outdoor

### Solution: SafeWalk

App that uses object detection and Visual-Large Language models to describe situations and give audio alerts

### Our Mission

Empower visually impaired individuals with the safety, independence, and confidence to navigate the world!

**Our Website**
https://zhangke626.github.io/safewalk-website/