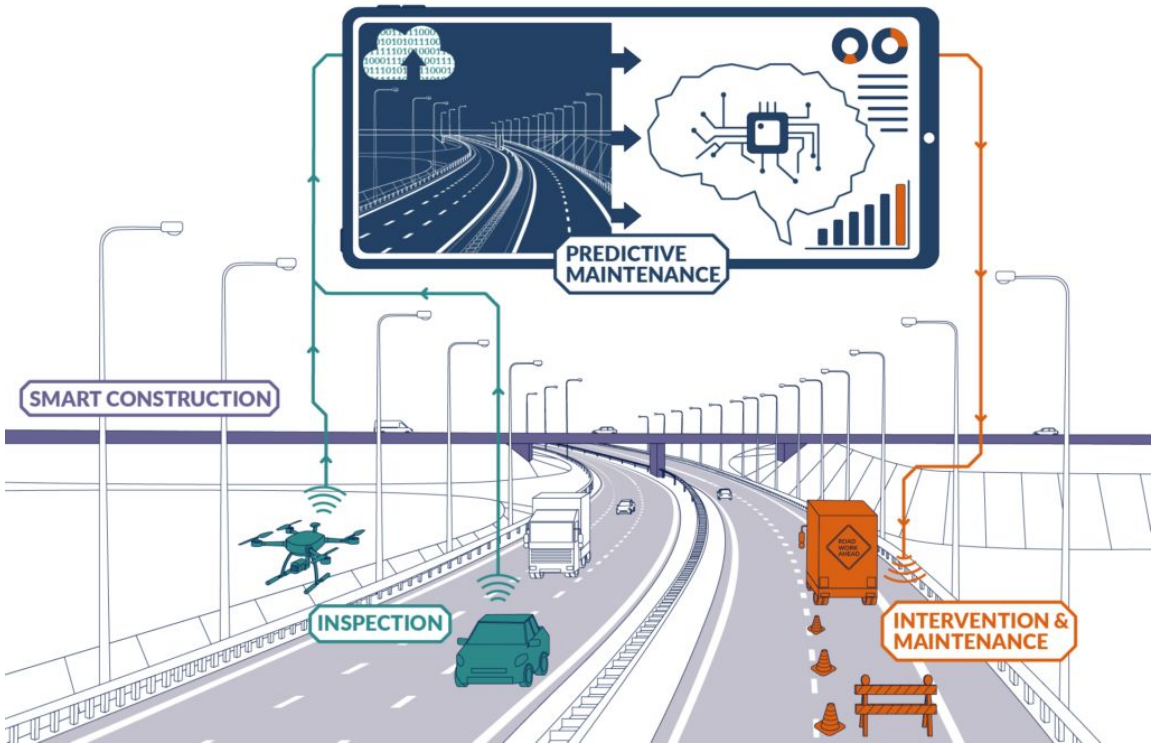
Enhanced YOLOv8-Based Object Detection of Road Assets Utilizing Generalized Focal Loss: A Case Study on **Thai Highway** Imagery

Teerapong Panboonyuen
https://kaopanboonyuen.github.io/

Natthasiri Rattanachona, Pawarisorn Thungthin, Nattapat Subsompon, Sittinun Thongbai, Weerachai Wongweeranimit, Ruttanawadee Phukham, and Teerapong Panboonyuen*

Enhanced YOLOv8-Based Object Detection of Road Assets Utilizing Generalized Focal Loss: A Case Study on Thai Highway Imagery

Overview of Object Detection of Road Assets

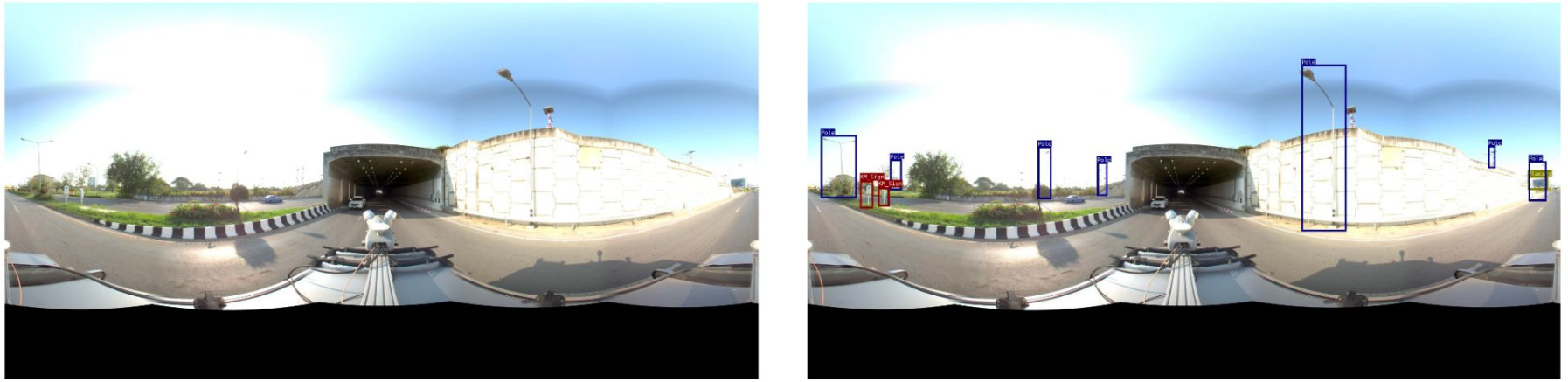


https://www.esci.eu/project/omicron-intelligent-road-asset-management-platform/

Motivation | "Why Road Asset Management Matters?"

Related Work

Panboonyuen, T.; Thongbai, S.; Wongweeranimit, W.; Santitamnont, P.; Suphan, K.; Charoenphon, C. Object Detection of Road Assets Using Transformer-Based YOLOX with Feature Pyramid Decoder on Thai Highway Panorama. Information 2022, 13, 5. https://doi.org/10.3390/info13010005

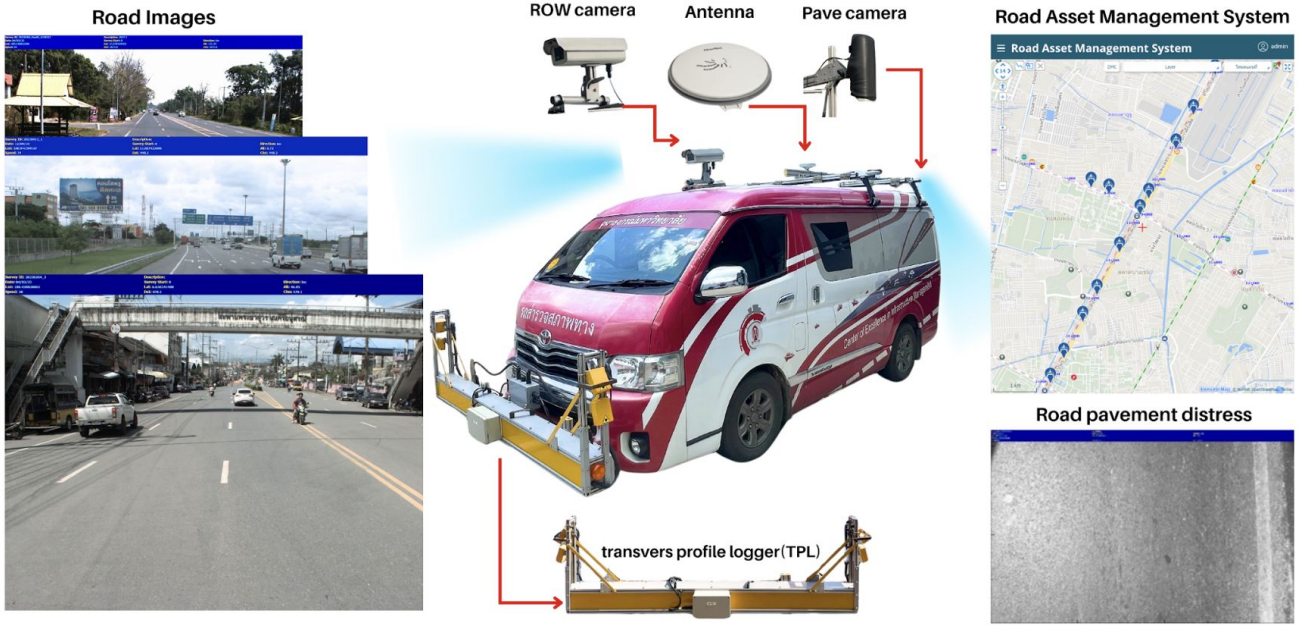


(a) (b)

Figure 1. The challenges in the Road Asset corpus. Sample of input image (**a**) and target image (**b**).

A Vehicle Equipped with Cameras (Our Contribution)

Self-Created Datasets: The Backbone of Our Approach



A Vehicle Equipped with Cameras (Our Contribution)

Self-Created Datasets: The Backbone of Our Approach

- The images were captured using the ZBR2-PGEHD-20S4C color camera, featuring a high-sensitive Sony ICX274 CCD sensor. Key specifications of the camera are as follows:
  - Model: ZBR2-PGEHD-20S4C Color 2.0 MP
  - Sensor: Sony ICX274 CCD, size 1/1.8
  - Shutter: Global Shutter
  - Resolution: 1624 x 1224 at 30 FPS (HD-SDI 25 FPS)
  - Interfaces: HD-SDI (2.97 Gbs/s) and GigE Vision (1000Mb/s, PoE)
  - Frame Rates: 2.0 MP at 30 FPS and 5.0 MP at 15 FPS

Dataset Overview

Total Images:

- **Training:** 1,994
- **Validation:** 1,418

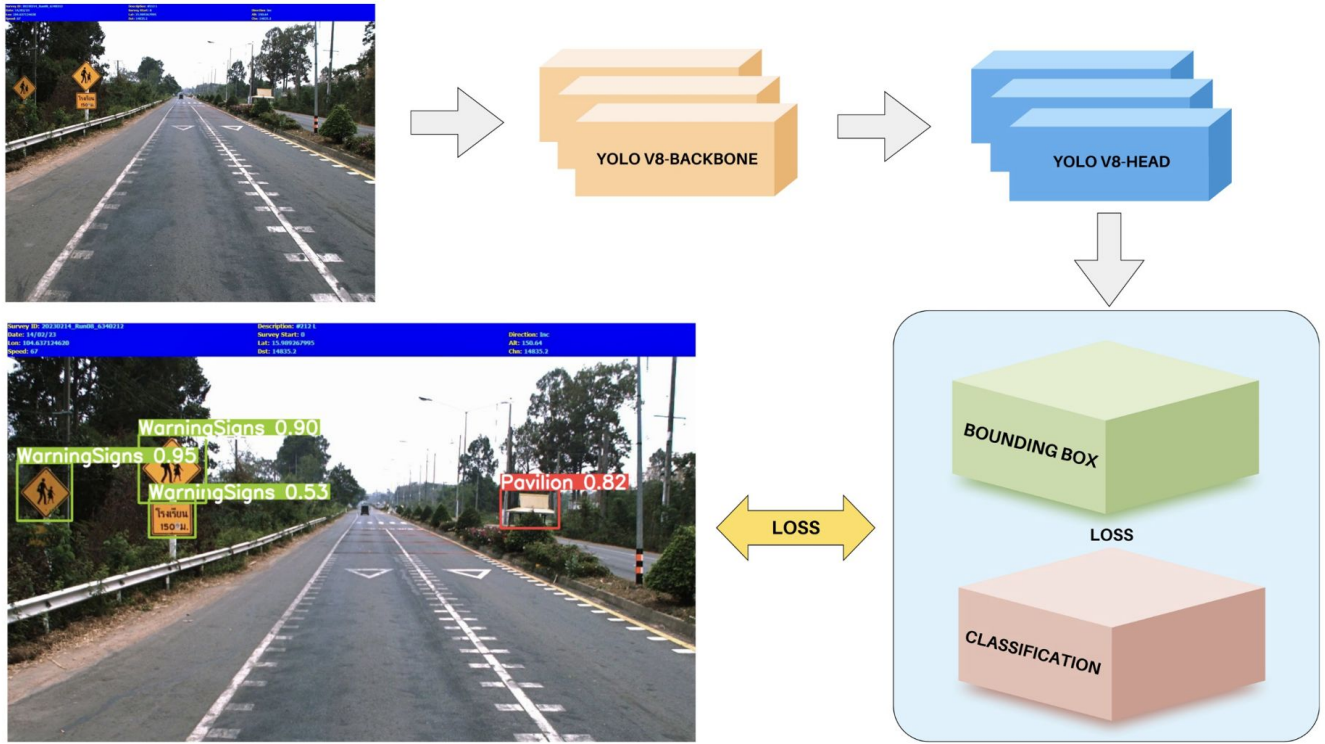
Resolution: 640x640 pixels
Source: Thai highway panoramas

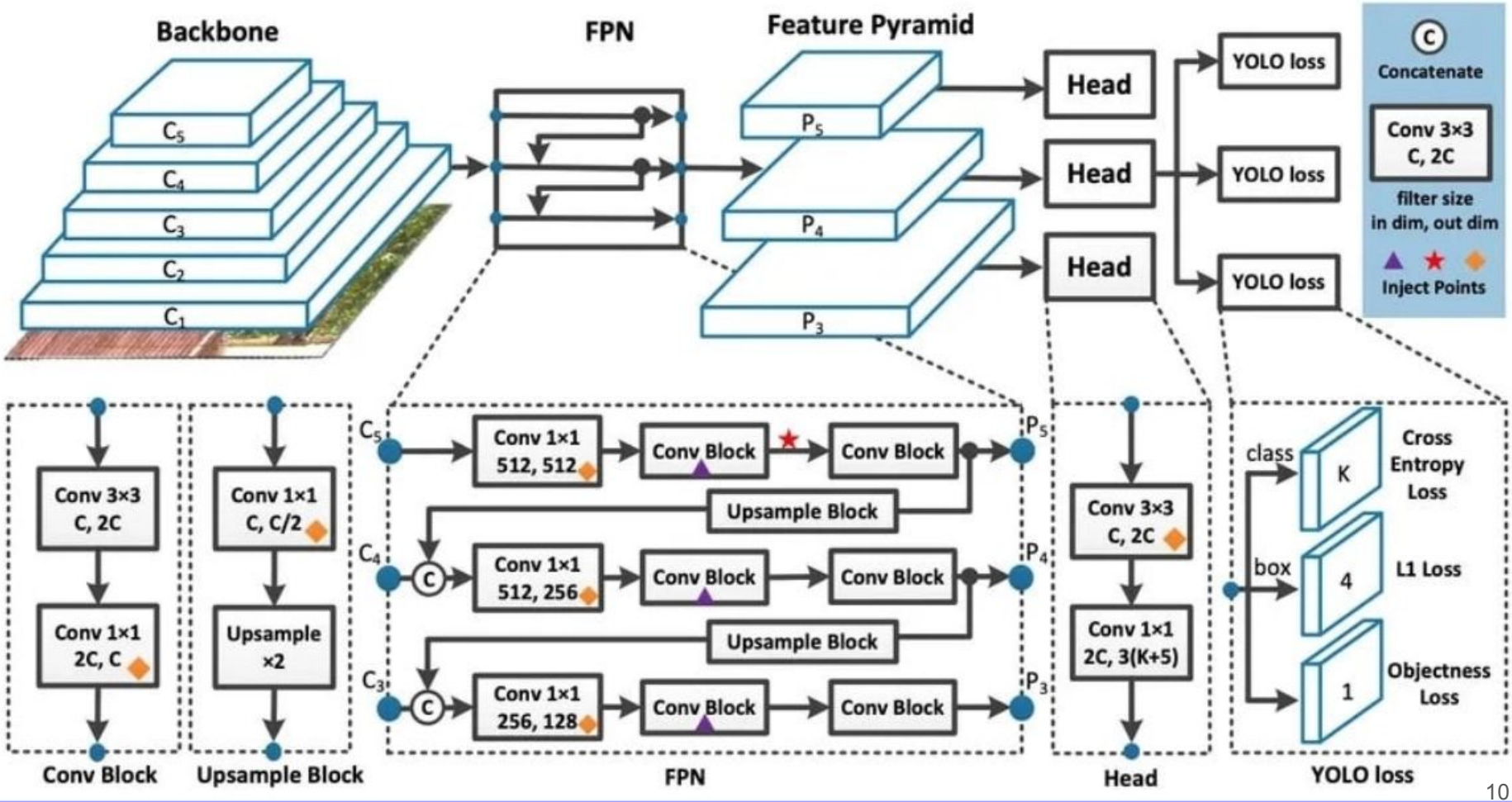
Capture Method: Mounted camera on Chulalongkorn University van

Annotations

- **Detection Tasks:** 7 classes
  - Pavilions
  - Pedestrian bridges
  - Information signs
  - Single-arm poles
  - Bus stops
  - Warning signs
  - Concrete guardrails
- **Segmentation Tasks:** 5 classes
  - Pavilions
  - Pedestrian bridges
  - Information signs
  - Warning signs
  - Concrete guardrails

Enhanced YOLOv8-Based Object Detection




Backbone
FPN
Feature Pyramid
C5
C4
C3
C2
C1
P5
P4
P3
Head
Head
Head
YOLO loss
YOLO loss
YOLO loss
C
Concatenate
Conv 3×3
C, 2C
filter size
in dim, out dim
Inject Points
Conv 3×3
C, 2C
Conv 1×1
2C, C
Conv Block
Conv 1×1
C, C/2
Upsample
×2
Upsample Block
C5
Conv 1×1
512, 512
Conv Block
Conv Block
P5
Upsample Block
C4
C
Conv 1×1
512, 256
Conv Block
Conv Block
P4
Upsample Block
C3
C
Conv 1×1
256, 128
Conv Block
Conv Block
P3
FPN
Conv 3×3
C, 2C
Conv 1×1
2C, 3(K+5)
Head
class
K
box
4
1
Cross
Entropy
Loss
L1 Loss
Objectness
Loss
YOLO loss

Generalized Focal Loss

Formula:

$$\mathrm{GFL}(p_t) = -\alpha_t (1 - p_t)^{\gamma_1} \log(p_t) - \beta_t p_t^{\gamma_2} \log(1 - p_t)$$

Why Generalized Focal Loss?

- **Adaptability**: Tailors focus on easy and hard examples by adjusting two parameters, γ_1 and γ_2.
- **Enhanced Class Imbalance Handling**: Provides better control over class imbalance compared to traditional Focal Loss.
- **Improved Detection Performance**: Balances precision and recall more effectively across diverse object detection scenarios.

Why Not Use Traditional Focal Loss?

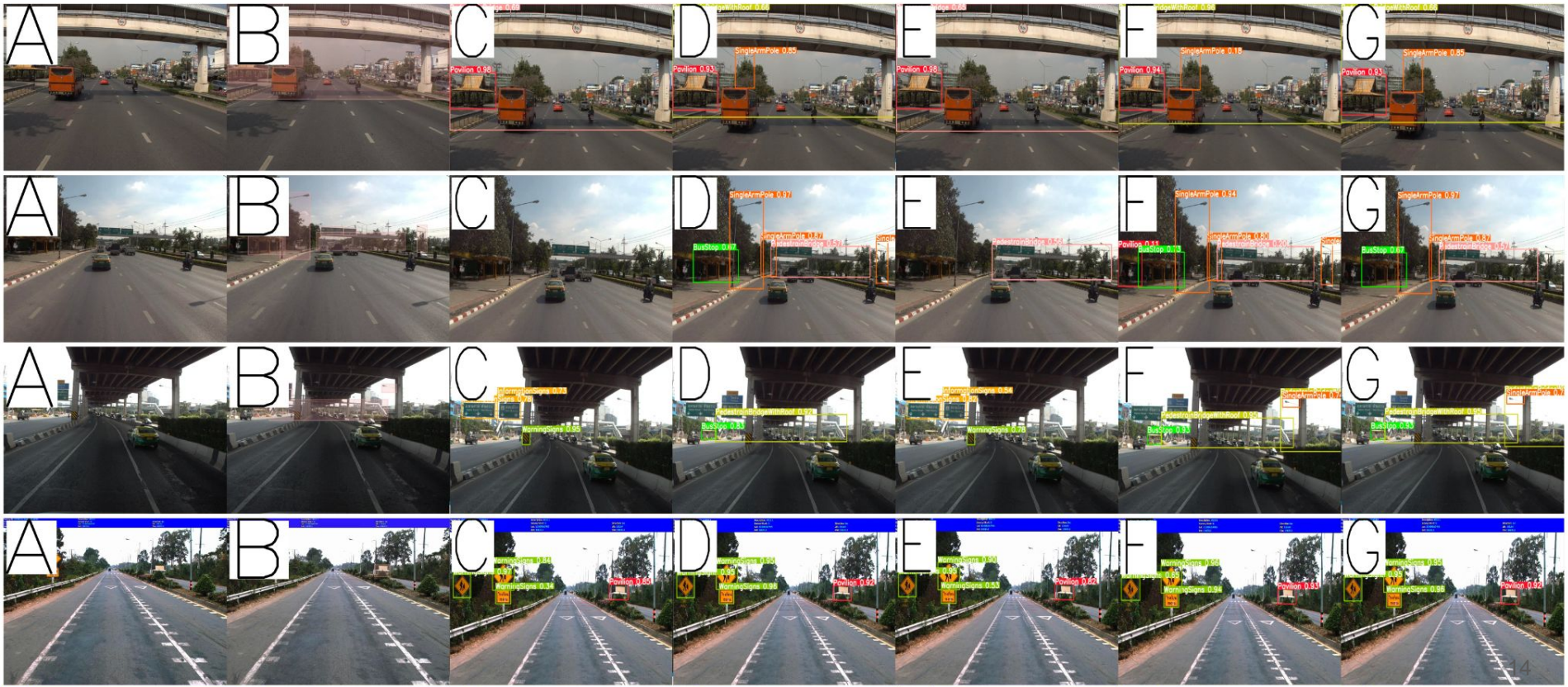
Limitations of Focal Loss:

- **Fixed Focus**: Focal Loss uses a single gamma parameter, which limits its ability to balance between easy and hard examples.
- **Rigid Handling of Class Imbalance**: It provides less flexibility in managing different degrees of class imbalance, which can affect detection performance.
- **Over-Suppression of Easy Examples**: Sometimes over-focuses on hard examples, leading to under-training on easy examples, which can reduce overall model accuracy.

Comparison of Detection Methods

Model	mAP50	mAP50-95	Precision	Recall	F1 Score
YOLOv8n	71.100	47.760	80.100	63.460	70.820
YOLOv8s	75.150	52.070	82.660	69.950	75.780
YOLOv8m	79.570	58.060	85.410	71.290	77.710
YOLOv8l	80.270	59.110	82.580	77.220	79.810
YOLOv8x	**80.340**	**60.840**	**79.100**	**76.680**	**77.870**

(a) Input, (b) Target, (c) Y OLOv8n, (d) Y OLOv8s, (e) Y OLOv8m, (f) Y OLOv8l, and (g) Y OLOv8x, with (f) and (g) representing our proposed method.



Comparison of Segmentation Methods (Masks)

Model	mAP50	mAP50-95	Precision	Recall	F1 Score
YOLOv8n	81.40	53.27	89.36	71.79	79.62
YOLOv8s	**90.08**	**63.88**	**88.78**	**86.09**	**87.41**
YOLOv8m	74.01	44.39	78.88	68.51	73.33
YOLOv8l	79.30	51.50	87.10	72.20	78.95
YOLOv8x	83.24	53.48	85.72	78.36	81.87

(a) Input, (b) Target, (c) Y OLOv8n, (d) Y OLOv8s, (e) Y OLOv8m, (f) Y OLOv8l, and (g) Y OLOv8x, with (f) and (g) representing our proposed method.

Conclusion and Q&A

- Enhanced detection with YOLOv8 and Generalized Focal Loss
- Significant improvements in accuracy and robustness
- Broad applicability and future potential
- Future research directions:
  - Real-time deployment
  - Further model optimizations