Comparing Object Detection, Instance Segmentation, and Semantic Segmentation for Automated Vegetation Detection in Railroad Systems

Mingyan Liu (Iris), Van Trung Le, Hwapyeong Song, Advay Chandramouli, Husnu S. Narman, and Ammar Alzarrad

IEEE UEMCon 2025

Marshall University NSF REU Program

01

Introduction

Motivation & Current Solution

140,000 miles

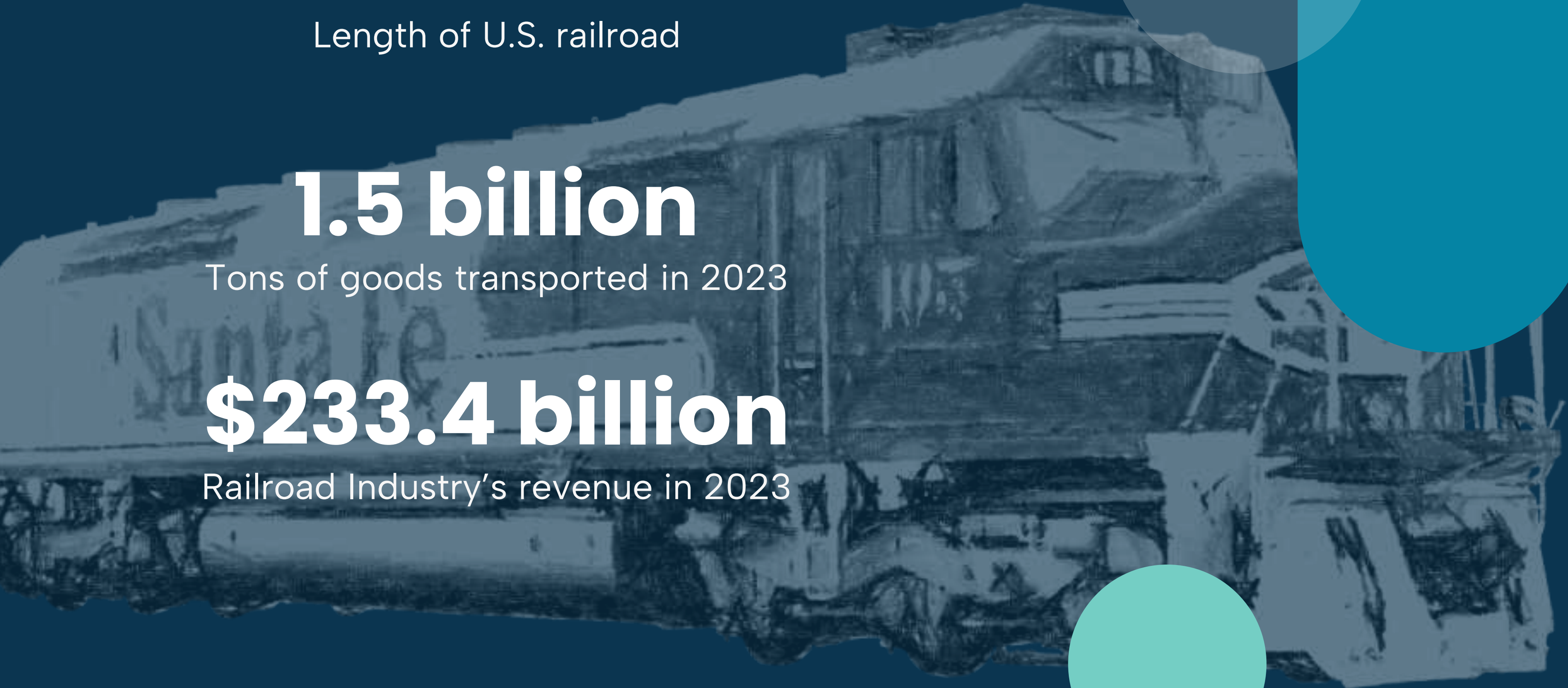
Length of U.S. railroad

1.5 billion

Tons of goods transported in 2023

$233.4 billion

Railroad Industry’s revenue in 2023



Consequences of Overgrown Vegetation

1
Fire Hazard

2
Slippery Rails

3
Track Deterioration

Current Solution

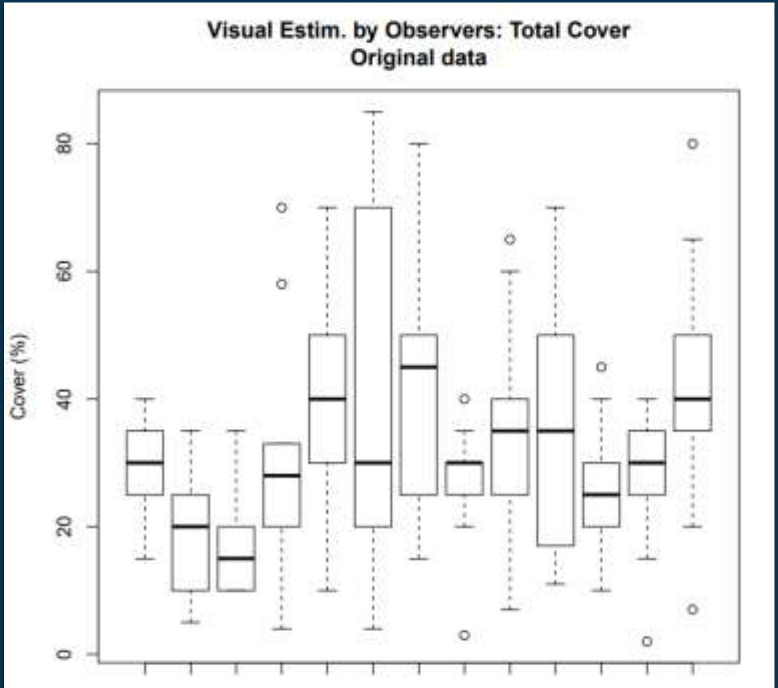
<u>Tradition methods of rail inspection</u>

Visual assessments conducted on site or through video footage

<u>This is a proven flawed method</u>

Nyberg et al. (2016)

Multiple ANOVA (Analysis of Variance) tests showed significant differences in mean rater estimates



Structural vs Organic Defects

02

Research Objective + Methodology

Compare modern deep learning model (YOLOv8, U-Net, DeepLabv3+) functions

- Compare YOLO Object Detection vs Instance Segmentation methods
- Compare U-Net & DeepLabv3+ Semantic Segmentation methods

Comparing domain specific vs general dataset

500 domain-specific vegetation dataset
vs 3,857 general vegetation dataset

YOLOv8: Object Detection vs. Instance Segmentation

Why YOLOv8? -> Fastest single-stage detector, with proven reliability in railroad defect real-time detection

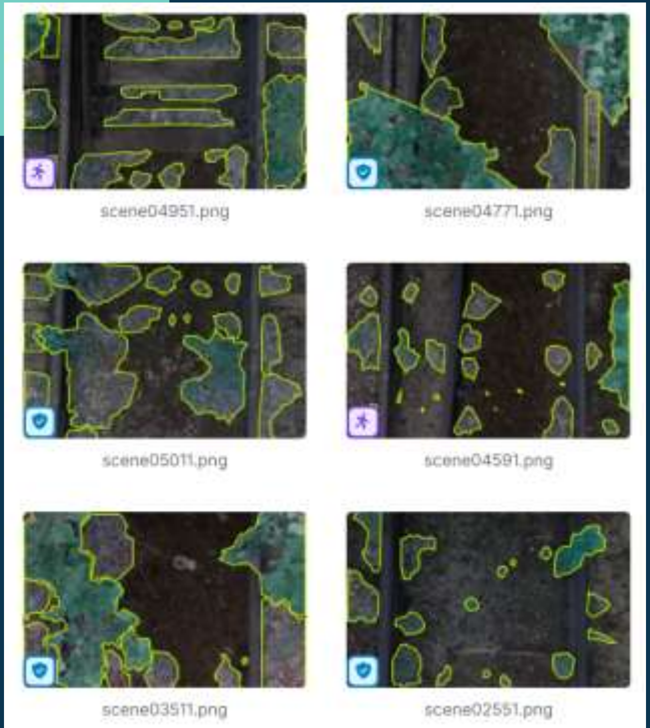
U-Net: Semantic Segmentation

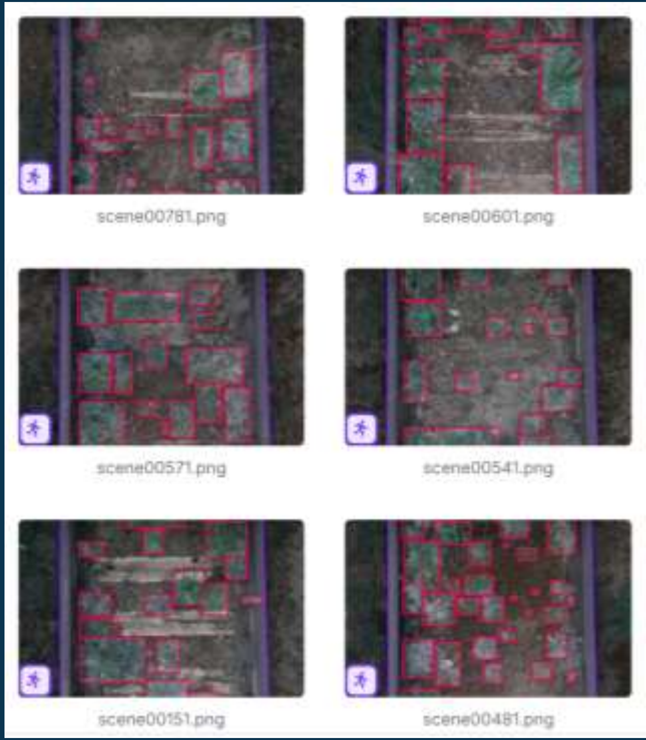
Why U-Net?-> U-shape encoder-decoder structure with skip connections preserves spatial detail during down sampling

DeepLabv3+: Semantic Segmentation

Why DeepLabv3+ -> U-shape encoder-decoder structure with skip connections preserves spatial detail during down sampling

Domain-Specific Dataset





500 railroad images at 5–15 mph using Intel RealSense D435, labeled in Roboflow (object + mask annotations)


4x4

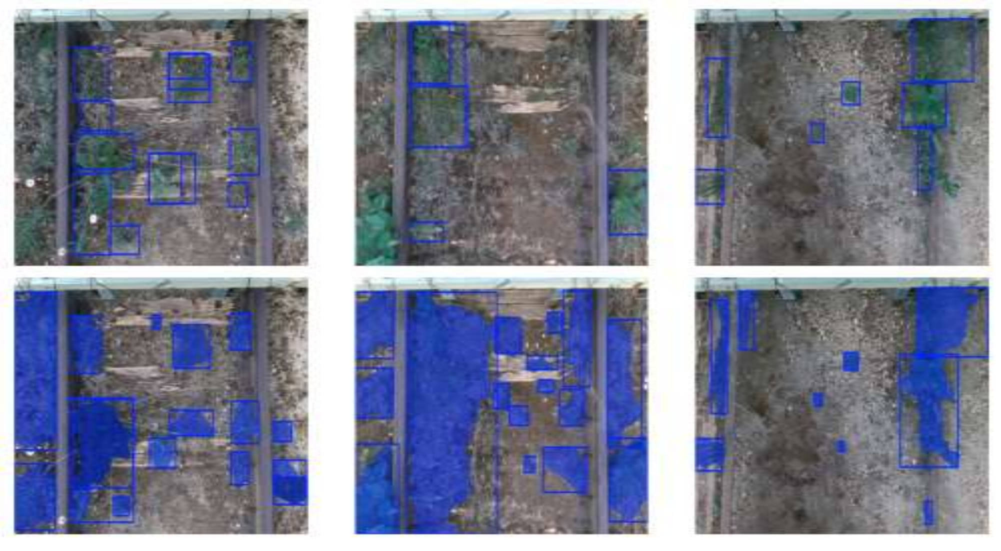
General Vegetation Dataset

On top of the original dataset, datasets forked from Roboflow was used.
9,865 total images after augmentation (rotation, noise, crop, zoom)

04

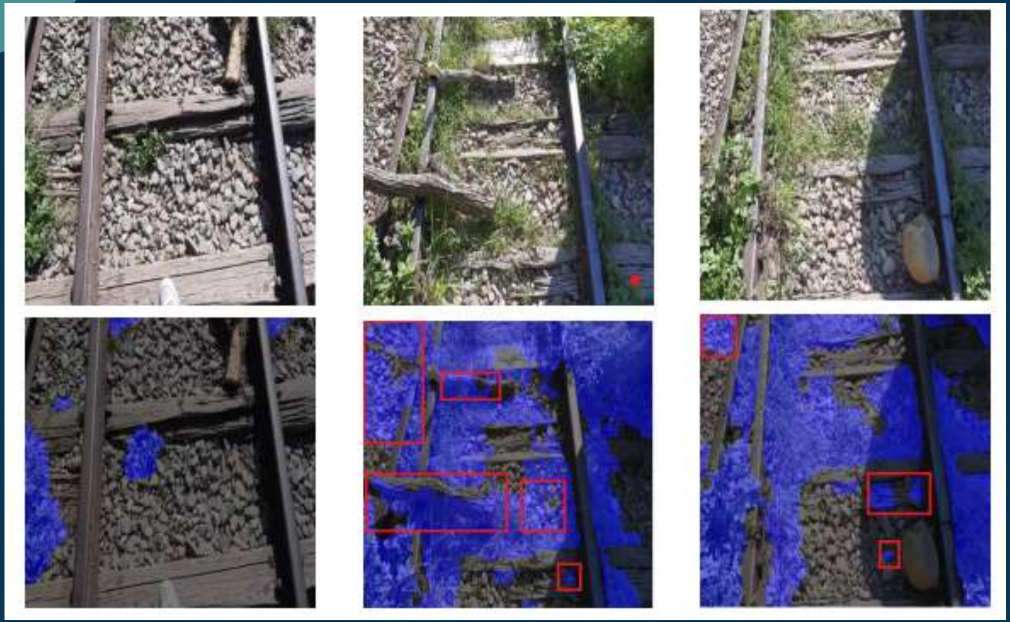
Results & Discussion

YOLOv8: Results



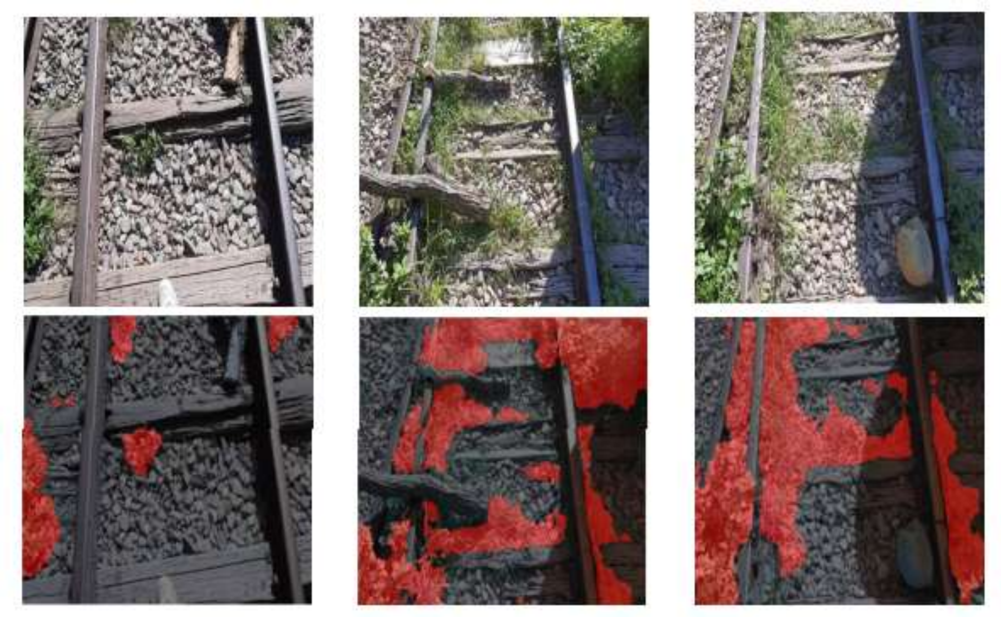
Metric	Object Detection	Segmentation
F1	0.69	0.72
Precision	1.00	1.00
Recall	0.88	0.86
mAP@0.5	0.68	0.73

U-net: Results



Metric	Value
Validation F1	0.8948
Validation Precision	0.9144
Validation Recall	0.8760
Validation IOU	0.8096
Validation Loss	0.1059

DeepLabv3+: Results



Metric	Value
Validation F1	0.9540
Validation Precision	0.9570
Validation Recall	0.9510
Validation IOU	0.9124
Validation Loss	0.053

Error Analysis

05

Conclusion

Summary

- DeepLabv3+ achieved the best overall metrics (F1 = 0.9540)

- YOLOv8 Segmentation performed better than object detection, but still weaker than semantic segmentation models

- Semantic segmentation is more suitable for irregular vegetation detection

YOLO Training Obstacles

Training Changes which made F-1 score decrease:

- Tuned hyparameters
  - IOU, Epochs, Learning Rate
- Data Augmentation
- Changing YOLO versions, weight sizes, types of optimizers
- Refining & Editing datasets
- Including 3-5% null images in dataset

/content/Vegetation-with-only-collected-data_SAM_Annotated-2-1	414 images with only field data (segmentation); no tuned parameters; normal	no tuned parameters	(mask F-1 score) 0.71	upgrade	train3
/content/Vegetation-with-only-collected-data_SAM_Annotated-2-1	414 images with only field data (segmentation); no tuned parameters; normal	optimizer=SGD	(mask F-1 score) 0.71	-	train 4
/content/Vegetation-with-only-collected-data_SAM_Annotated-2-1	414 images with only field data (segmentation); no tuned parameters; normal	optimizer=AdamW	(mask F-1 score) 0.72	upgrade	train5
/content/Vegetation-with-only-collected-data_SAM_Annotated-2-1	414 images with only field data (segmentation); no tuned parameters; normal	yolov11n-seg	(mask F-1 score) 0.72	-	train6
/content/Vegetation-with-only-collected-data_SAM_Annotated-2-1	414 images with only field data (segmentation); no tuned parameters; medium	yolov8 n to m (weight size)	(mask F-1 score) 0.71	downgrade	train7

Future Work Considerations

- Integrate binary railway masks for selecting ROI

- Train YOLOv8 using general vegetation dataset & UNet/DeepLabv3+ using domain specific dataset

- Evaluating model variants

Thank You!

This research was supported by the NSF REU Program for Data Analytics at Marshall University.

Special thanks to my research mentor Dr. Husnu Narman, and program chairs Dr. Haroon Malik and Dr. Yousef Fazea.