Bounding-box to Segmentation Labeling for Pavement Distresses Using a Closed-loop Platform on 2D/3D Images

Wenhan Tao[1], Tanzina Akter Tani[1], Jelena Tešić[1], Feng Wang[2], Yongsheng Bai[2]

[1] College of Science and Engineering and [2] Ingram School of Engineering, Texas State University

Background and Objective

Motivation
Bounding boxes are easy to obtain but cannot capture precise distress boundaries. Segmentation labels provide pixel-level detail but are costly to annotate manually. This makes large-scale, high-quality mask annotation difficult in practice for pavement distress analysis.

Objective
Develop a practical workflow to convert existing bounding-box annotations into segmentation labels while reducing manual effort.

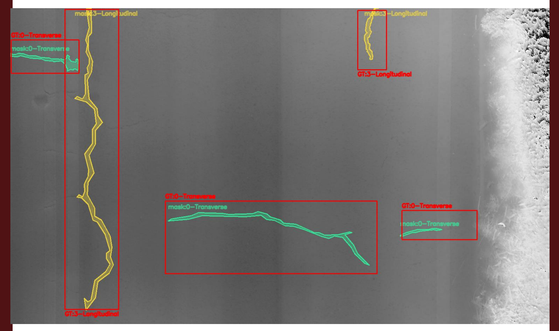

Figure 1. Example bounding boxes and segmentation masks for pavement distress.

Dataset Description

The experiments used the 2024 TxDOT Jointed Concrete Pavement (JCP) dataset. It includes aligned 2D intensity images and 3D range measurements collected by a pavement inspection vehicle. The road surface was divided into 1536 × 900 tiles for annotation and training. The dataset contains multiple distress classes with strong class imbalance, including rare categories.

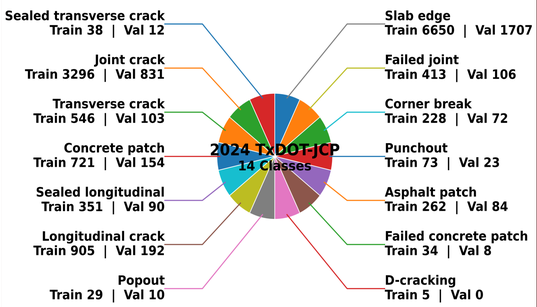

Figure 2. Overview of class distribution in JCP dataset.

Proposed Method

Closed-loop Box-to-Segmentation Workflow
We propose a closed-loop workflow that converts existing bounding-box labels into segmentation annotations. SAM3 first generates initial masks within box regions, and annotators refine them in local CVAT. The refined labels are then used to train a YOLOv8-s segmentation model to predict segmentation masks.

Step 1. Box-guided mask initialization in CVAT.
Bounding boxes and initial SAM3 masks are imported into Nuclio-connected local CVAT, enabling secure SAM3-assisted annotation with box location and class references.

Step 2. Dual-modality human refinement.
Intensity and range images are organized as two aligned CVAT tasks, enabling annotators to refine masks using the modality in which each distress is more clearly visible.

Step 3. Rule-based label export and merging.
After refinement, labels from the two tasks are exported and merged under a strict rule: each distress instance is kept in only one modality to avoid duplication.

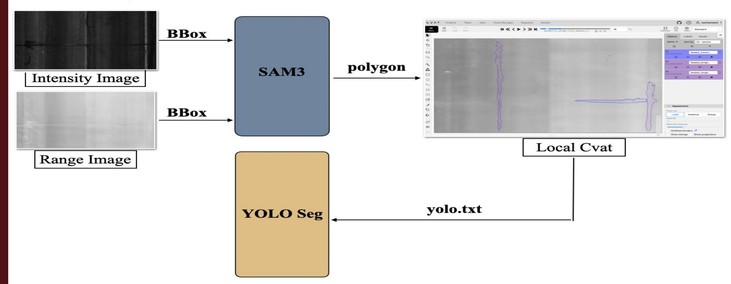

Figure 3. Main closed-loop box-to-segmentation workflow.

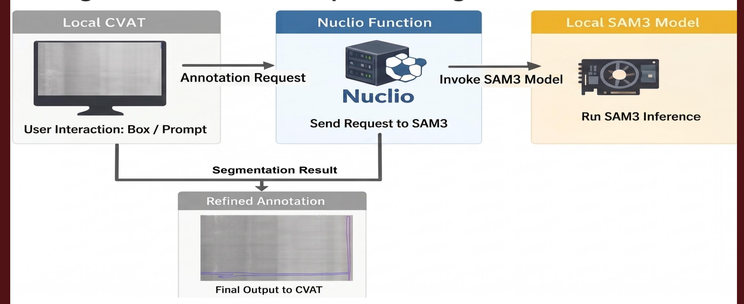

Figure 4. Local CVAT–SAM3 integration for interactive annotation.

Experimental Setup

All models were trained for 150 epochs with an input size of 640.
Both detection and segmentation used fused intensity–range images.

$$\text{Precision} = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}$$

$$\text{mAP@0.5} = \textit{mean Average Precision at IoU} = 0.5$$

Annotation Protocol

Annotation Time Recording
Annotation time was recorded during real labeling sessions by two trained annotators. On average, each annotator refined about 140 images, covering more than 300 distress instances in total. Because early records were not separated by class, the measured time reflects both distress searching and mask refinement effort.

Pseudo-Mask Quality Levels
SAM3 pseudo masks were grouped into high, medium, and low quality based on practical usefulness for annotation, considering completeness, thin-crack preservation, coverage accuracy, and over-coverage.
High-quality: about 80% time reduction
Medium-quality: about 30–50% reduction
Low-quality: little or no practical benefit.

Table 1. Annotation time for two rare classes

Class	Manual Annotation	SAM3-Assisted Refinement	Time Reduction
Failed concrete patch	3–5 min	2–3 min	~30–40%
Popout	1–2 min	~30 s	~50–75%

Results and Discussion

Using the original bounding-box annotations, the YOLOv8-s box model gave limited performance on the 2024 TxDOT JCP dataset, especially for rare classes, suggesting that box labels were insufficient to capture fine distress details and textures. The proposed workflow efficiently converted box annotations into refined segmentation labels through SAM3-assisted initialization and human correction. Training YOLOv8-s segmentation on these refined labels led to substantially better results, as summarized in Table 2.

Table 2. AP@0.5 comparison for rare classes and overall performance

Class	Baseline	Refined (Seg)	Gain
Failed concrete patch	0.464	0.761	+0.297
Popout	0.007	0.723	+0.716
All classes	0.235	0.742	+0.507

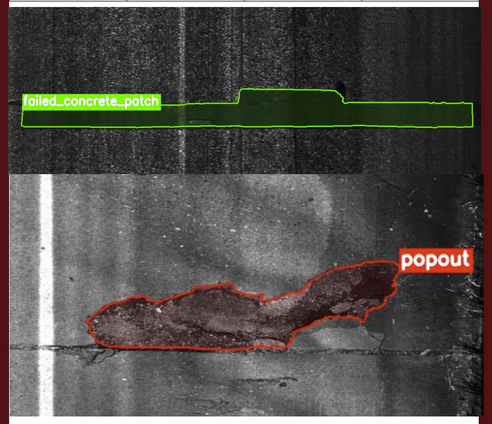

Figure 5. Rare-class prediction examples

Conclusion and Future Work

We presented a practical box-to-segmentation workflow using SAM3, local CVAT, and human refinement. The workflow reduced labeling effort and produced useful segmentation labels for training. Results on rare classes showed clear improvement over the original box-based baseline. Future work will expand the annotated dataset and develop a task-specific assistant for automatic mask initialization.

Acknowledgment

This research was also sponsored by the Texas Department of Transportation (project No. 0-7150).

RPUG Road Profile Users' Group

TEXAS STATE UNIVERSITY®
MEMBER THE TEXAS STATE UNIVERSITY SYSTEM