Cascaded Workflow for Automated Pavement Crack Detection and Measurement Using Generative AI Models and Image Processing Techniques

Yongsheng Bail and Feng Wang
Ingram School of Engineering, Texas State University

Abstract

In this research, Generative AI (Artificial Intelligence) models such as Ground DINO and Florence 2 are developed in the mission of automated pavement distress detection with 2D/3D image data, and computer vision techniques are applied for the quantification of pavement cracks. In this cascaded workflow, the Generative AI models are compared with different pavement distress detection models and developed with the dataset collected and annotated by the research team. After the surface defects are detected, a technique using the Structured Forest edge extraction method is proposed to delineate the shape of the cracks and then measure the crack geo-information (widths, lengths, and area) at the distress locations. Finally, weighted cracking widths, lengths, and areas are applied to calculate the distress severity levels or distress scores to follow the protocols of distress identification guidelines and pavement management standards specified by the U.S. and state DOTs. This two-step workflow with the Generative AI and image processing techniques will provide an automated pipeline to process the field collected pavement image data and assess the cracking condition of the pavement surface with cost-effectiveness.

Introduction

1. Traditional and open-vocabulary object detection

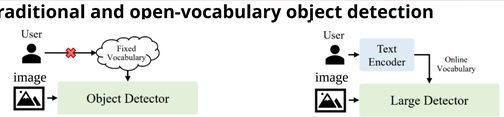

Figure 1 Fixed-vocabulary (a) and open-vocabulary (b) detection.

2. Models used for pavement distress detection

1) Real-Time Detection Transformer (RT-DETR)

Images are the input. Intrascale feature interaction (AIFI) and cross-scale feature-fusion module (CCFM) are the encoder. IoU-aware query is used for image feature selection. The decoder generating boxes and confidence scores .

2) YOLOv12

YOLOv12 is an attention-centric variant of the YOLO (You Only Look Once) family.

3) YOLO World

Text and the input image are encoded with YOLO backbone.

4) Grounding-DINO

With ground pre-training, the Grounding DINO (Distillation with No Labels) can utilize a feature enhancer, a language-guided query selection and a modality fusion with a cross-modality decoder.

5) Florence 2:

The semantic granularity at three stages forms a comprehensive annotation module named FLD-5B. Florence-2 is a unified architecture based on FLD-5B.

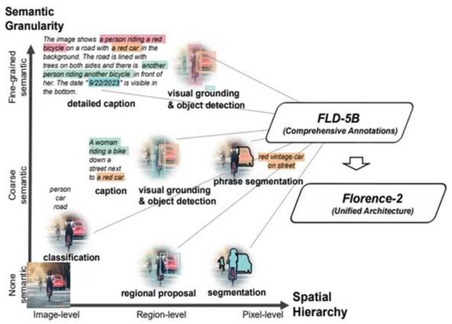

Figure 2. An illustration of Florence 2 enabling semantic granularity and spatial hierarchy as a vision foundation model.

Methodology

1. Data collection

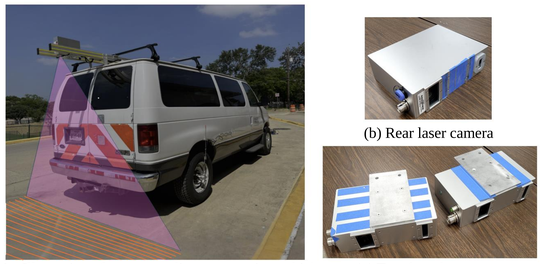

Figure 3 2D/3D laser pavement distress scanning vehicle and its sensors.

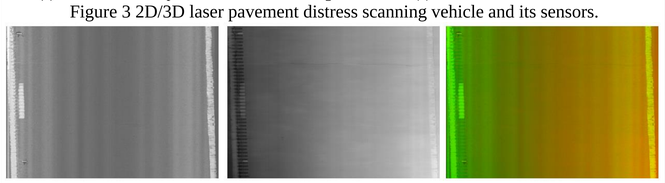

Figure 4 An example of 2D, 3D, and fused images for the collected data.

2. Cascaded workflow

1) Pavement distress detection with Generative AI models.
2) Crack segmentation with the Structured Forest edge detection method.

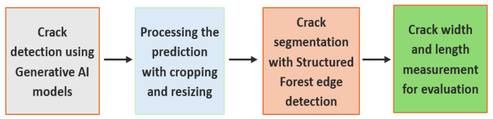

Figure 5 Flowchart of using Generative AI models and Structured Forest edge detection for pavement crack measurements.

Step 1: Data preparation

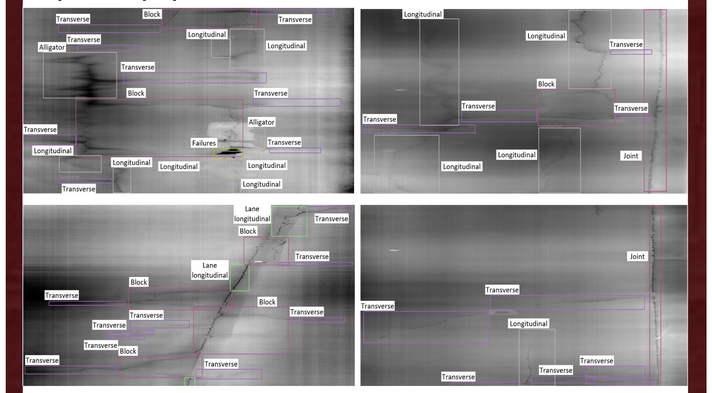

Figure 6 Four examples of pavement distress labeled by us for Asphalt Concrete Pavement (ACP).

Step 2: Distress detection and cropped with Generative AI models

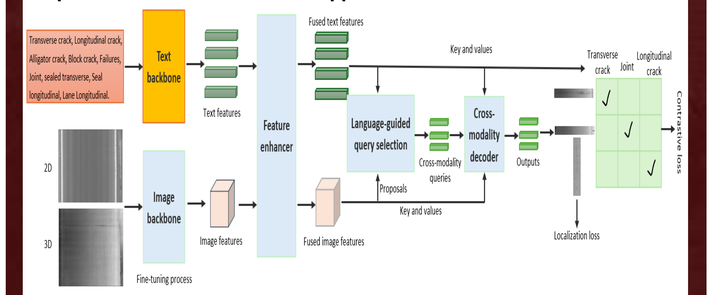

Figure 7 Generative AI models (vision-language models) for pavement distress detection.

Step 3: Crack measurements

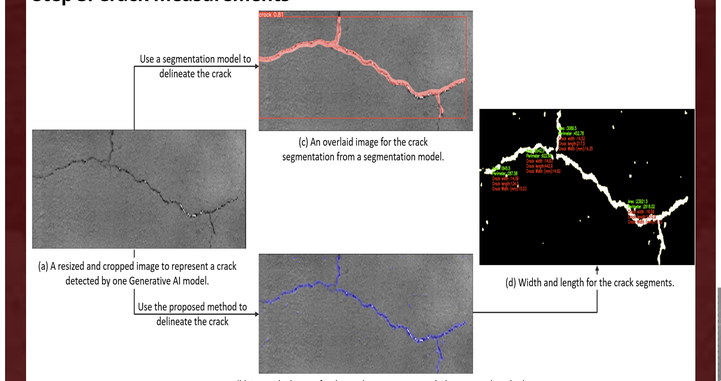

Figure 8 Crack measurements with a cropped image. (a) is the image detected with a piece of crack after resizing and cropping. (b) is the overlaid image of crack segmentation with Structured Forest edge detection method. (c) is the result of using a crack segmentation neural network. (d) is the width and length information for each crack segment.

Results

1. Data preprocessing and augmentation

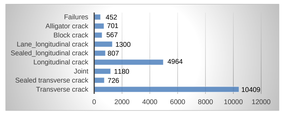

Figure 9 Distribution of different ACP pavement distresses for the original training dataset.

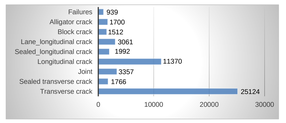

Figure 10 Distribution of different ACP pavement distresses for the tiled training dataset.

2. Model performance with and without data augmentation

Table 1. Results for RT-DETR model trained by fused images with and without tiling

Class name	Without tiling			With tiling		
	Precision	Recall	mAP50	Precision	Recall	mAP50
Transverse crack	0.344	0.482	0.354	0.663	0.602	0.632
Sealed transverse crack	0.517	0.254	0.296	0.567	0.516	0.53
Joint	0.204	0.669	0.421	0.82	0.815	0.847
Longitudinal crack	0.325	0.253	0.197	0.539	0.481	0.459
Sealed longitudinal crack	0.257	0.349	0.245	0.657	0.573	0.589
Lane longitudinal crack	0.258	0.289	0.145	0.588	0.551	0.518
Block crack	0.0486	0.0339	0.0184	0.371	0.285	0.249
Alligator crack	0.0991	0.0118	0.0603	0.614	0.509	0.538
Failures	0.376	0.268	0.289	0.621	0.651	0.663
Overall performance	0.27	0.29	0.225	0.604	0.553	0.558

3. Model performance with 3D and 2D/3D fused images

Table 2. Results for RT-DETR model trained by tiled 3D and fused 2D/3D images

Class name	3D only			Fused 2D/3D		
	Precision	Recall	mAP50	Precision	Recall	mAP50
Transverse crack	0.647	0.615	0.641	0.663	0.602	0.632
Sealed transverse crack	0.568	0.54	0.506	0.567	0.516	0.53
Joint	0.824	0.821	0.859	0.82	0.815	0.847
Longitudinal crack	0.544	0.488	0.485	0.539	0.481	0.459
Sealed longitudinal crack	0.622	0.541	0.582	0.657	0.573	0.589
Lane longitudinal crack	0.556	0.514	0.478	0.588	0.551	0.518
Block crack	0.355	0.259	0.225	0.371	0.285	0.249
Alligator crack	0.584	0.463	0.507	0.614	0.509	0.538
Failures	0.514	0.687	0.624	0.621	0.651	0.663
Overall performance	0.579	0.548	0.545	0.604	0.553	0.558

4. Performance of traditional and Generative AI models with fused 2D/3D images

Table 3. Overall performance of different models

Model name	3D or fused	Data augmentation	mAP50	Precision	Recall
RT_DETR	3D	yes	0.545	0.579	0.548
RT_DETR	fused	yes	**0.558**	0.604	**0.553**
YOLOv12	fused	yes	0.546	**0.610**	0.526
YOLO World	fused	yes	0.511	0.570	0.516
Grounding-DINO	fused	yes	0.515	0.522	0.504
Florence 2	3D	yes	0.301	0.400	0.440
Florence 2	fused	yes	0.263	0.450	0.360

5. An example of crack measurements

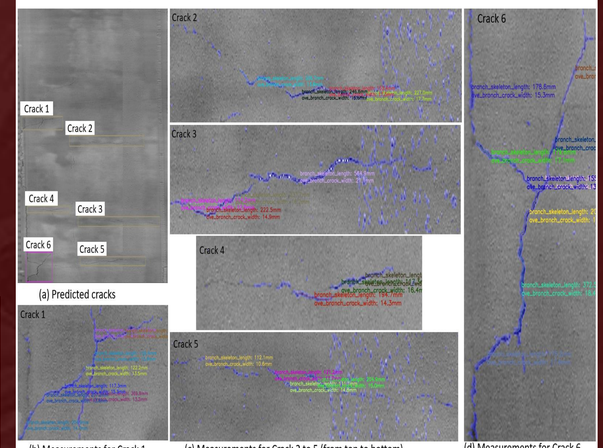

Figure 11 An illustration of measuring cracks detected by the Grounding DINO model. (a) is the predicted cracks from the model. (b) to (d) are the length and width of each crack branch (segment). The measurements have the same colors as the branch skeletons of cracks on the overlaid images.

Conclusions

Several findings are observed through this research:

1) Generative AI models can be used for pavement distress detection, but traditional deep learning methods like RT-DETR can compete with them.

2) An imbalance problem for some types of pavement distresses in the training dataset affects all the models. It is still insufficient after using data augmentation techniques to increase the dataset.

3) Structured Forest edge detection on the crack segmentation works well for the proposed framework, but there are cases that this method fails to capture the edges of thin or fading cracks.

4) Crack measurements need to be verified with established data libraries and provide statistical results for evaluation.

5) In the future, segmentation networks will be applied on crack delineation to get more accurate crack segmentation on ACP.

6) Crack qualification on JCP and CRCP will be the next research.

Acknowledgements

This research is part of the project "using artificial intelligence to improve the accuracy of automated pavement condition data collection" , funded by the U.S. National Science Foundation (grant No. 2213694).

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