Coastal REsearch And Transportation Education
Tier 1 University Transportation Center
U.S. Department of Transportation

Deep Learning Based Automated Data Collection Technology for Coastal Highway Pavements

August 31, 2025

(1) Feng Wang, Professor, Texas State University, Ingram School of Engineering, 601 University Drive, San Marcos, TX 78666; ORCID https://orcid.org/0000-0002-1528-9711; Email: f_w34@txstate.edu
(2) Carlos Sanchez, Graduate Research Assistant, Texas State University, Ingram School of Engineering, 601 University Drive, San Marcos, TX 78666; ORCID https://orcid.org/ 0009-0003-6713-0483; Email: cds242@txstate.edu
(3) Yongsheng Bai, Postdoc Research Fellow, Texas State University, Ingram School of Engineering, 601 University Drive, San Marcos, TX 78666; ORCID https://orcid.org/0000-0002-2762-8709; Email: ysh_bai@txstate.edu
(4) Xiaohua Luo, Assistant Professor of Instruction, Texas State University, Ingram School of Engineering, 601 University Drive, San Marcos, TX 78666; ORCID https://orcid.org/0000-0002-0686-450X; Email: xiaohualuo@txstate.edu

Final Research Report
Prepared for:
Coastal REsearch And Transportation Education

CREATE

Technical Report Documentation Form

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
2301-6	01895142	n/a

4. Title and Subtitle

Deep Learning Based Automated Data Collection Technology for Coastal Highway Pavements

5. Report Date

8/31/2025

6. Performing Organization Code

n/a

7. Authoris: Wang Feng, Professor, Texas State University, ISOE 0000-0002-1528-9711 email: f_w34@txstate.edu; Carlos Sanchez, Graduate Research Assist. Texas State University
0009-0003-6713-0483 email: cds242@txstate.edu; Yonsheng Bai, Post Doc Research Fellow, Texas State
0000-0002-2762-8709 email:ysh_bai@txstate.edu; Xiaohua Luo, Assist Professor of Instructions TXST

8. Performing Organization Report No.

n/a

9. Performing Organization Name and Address

Ingram School of Engineering, Texas State University, 601 University Dr. San Marcos, TX 78666

10. Work Unit No. (TRAIS)

n/a

12. Sponsoring Agency Name and Address

Office of the Assistant Secretary for Research and Technology
University Transportation Centers Program
Department of Transportation
Washington, DC United States 20590

11. Contract or Grant No.

69A3552348330

13. Type of Report and Period Covered

Final Project Report
09/01/2023-8/31/2025

14. Sponsoring Agency Code

OST-R

15. Supplementary Notes

https://create.engineering.txst.edu/

16. Abstract

With the recent federal legislation approval of the funding of the United States's infrastructure, investment into transportation research has prompted the implementation of new technologies to advance the data collection and evaluation methods for transportation asset management. As departments of transportation (DOTs) upgrade their equipment to utilize these technologies, a higher standard for the quality of highway infrastructure across the states has raised concerns for an improvement in traffic safety and efficient maintenance and rehabilitation of the roadways. This is essential for pavements in locations where surface distresses are developed at a higher rate, such as coastal regions due to rapid development and extreme weather phenomena being more frequent and intensive. State DOTs have adopted the use of automated pavement data collection and condition evaluation. However, the inadequate accuracy of the existing automated technology has led to erroneous distress measurements and inconsistent manual intervention approaches for pavement engineers to assess the surface damage of the pavements after data collection to establish more reliable analyses of the performances. With the recent advancements in artificial intelligence and deep learning, progress towards more accurate and efficient detection algorithms has provided possibilities of better data quality to monitor the surface condition of the pavements. Due to these new technologies, 2D/3D pavement images can be analyzed with improved accuracy for more efficient detection of pavement distresses such as cracks caused by various factors like environment, weather, age, and traffic loading. In this study, 2D/3D images were collected from pavements in the coastal regions in the states of Louisiana, Mississippi, and Texas. The proximity of these states to the Gulf of Mexico and the presence of rapid population growth and economic development in this region, unique distress image data were acquired to train and evaluate models such as convolution neural networks (CNN) and a visual transformer after manual annotations were made. This research makes two contributions: the establishment of a novel multi-modal pavement distress dataset utilizing the combination of high-resolution 2D/3D imagery across inland and coastal regions and a systematic evaluation of state-of-the-art detection models. The images consisted of high quality 4096x2048 pixel resolution scans that were 47 feet in the longitudinal direction and 14 feet in the transverse direction on both asphalt and concrete surfaces. Deep learning models such as the You Only Look Once (YOLO) line of detection models and Real-Time Detection Transformer (RT-DETR) were developed by training, validation, and testing on the pavement image data collected in this study to more efficiently evaluate the pavement surface condition of the coastal regions. Rigorous labelling and revisions within the dataset were conducted to further optimize the detection accuracy of models as selection continued. Comparisons between the inland and coastal data in asphalt pavements were conducted to determine model capabilities for expansion as the sections near the Gulf coast had lacked sufficient instances of a few classes. Detection capabilities of the selected model faced many issues that would be reflected in real-world conditions, such as noise, distress severity, distress density, and abnormal distress appearances. These factors would heavily affect the accuracy for each distress, with joints on asphalt surfaces reaching a peak mAP50 of 0.928 while asphalt patches would reach a peak mAP50 of 0.981. Despite retaining lower scores on some classes, this study displays the effectiveness of deep learning detection models on pavement datasets with distresses containing a large variety of distresses and common objects.

17. Key Words	18. Distribution Statement
pavement, automated, distress detection, artificial intelligence, deep learning, accuracy	No Restrictions

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	71	n/a

ACKNOWLEDGMENT

This study was funded, partially or entirely, by the U.S. Department of Transportation through the Coastal REsearch and Transportation Education University Transportation Center under Grant Award Number 69A3552348330. The work was conducted at Texas State University. The authors thank project advisory board member Dr. Robin Huang. The authors would also like to thank the Texas Department of Transportation (TxDOT) and National Science Foundation for the assistance in/with this research.

DISCLAIMER

ABSTRACT

With the recent federal legislation approval of the funding of the United States's infrastructure, investment into transportation research has prompted the implementation of new technologies to advance the data collection and evaluation methods for transportation asset management. As departments of transportations (DOTs) upgrade their equipment to utilize these technologies, a higher standard for the quality of highway infrastructure across the states has raised concerns for an improvement in traffic safety and efficient maintenance and rehabilitation of the roadways. This is essential for pavements in locations where surface distresses are developed at a higher rate, such as coastal regions due to rapid development and extreme weather phenomena being more frequent and intensive. State DOTs have adopted the use of automated pavement data collection and condition evaluation. However, the inadequate accuracy of the existing automated technology has led to erroneous distress measurements and inconsistent manual intervention approaches for pavement engineers to assess the surface damage of the pavements after data collection to establish more reliable analyses of the performances. With the recent advancements in artificial intelligence and deep learning, progress towards more accurate and efficient detection algorithms has provided possibilities of better data quality to monitor the surface condition of the pavements. Due to these new technologies, 2D/3D pavement images can be analyzed with improved accuracy for more efficient detection of pavement distresses such as cracks caused by various factors like environment, weather, age, and traffic loading. In this study, 2D/3D images were collected from pavements in the coastal regions in the states of Louisiana, Mississippi, and Texas. The proximity of these states to the Gulf of Mexico and the presence of rapid population growth and economic development in this region, unique distress image data were acquired to train and evaluate models such as convolution neural networks (CNN) and a visual transformer after manual annotations were made. This research makes two contributions: the establishment of a novel multi-modal pavement distress dataset utilizing the combination of high-resolution 2D/3D imagery across inland and coastal regions and a systematic evaluation of state-of-the-art detection models. The images consisted of high quality 4096x2048 pixel resolution scans that were 47 feet in the longitudinal direction and 14 feet in the transverse direction on both asphalt and concrete surfaces. Deep learning models such as the You Only Look Once (YOLO) line of detection models and Real-Time Detection Transformer (RT-DETR) were developed by training, validation, and testing on the pavement image data collected in this study to more efficiently evaluate the

pavement surface condition of the coastal regions. Rigorous labelling and revisions within the dataset were conducted to further optimize the detection accuracy of models as selection continued. Comparisons between the inland and coastal data in asphalt pavements were conducted to determine model capabilities for expansion as the sections near the Gulf coast had lacked sufficient instances of a few classes. Detection capabilities of the selected model faced many issues that would be reflected in real-world conditions, such as noise, distress severity, distress density, and abnormal distress appearances. These factors would heavily affect the accuracy for each distress, with joints on asphalt surfaces reaching a peak mAP50 of 0.928 while asphalt patches would reach a peak mAP50 of 0.981. Despite retaining lower scores on some classes, this study displays the effectiveness of deep learning detection models on pavement datasets with distresses containing a large variety of distresses and common objects.

Keywords: pavement, automated, distress detection, artificial intelligence, deep learning, accuracy

Table Of Contents

List of Figures

List of Tables

INTRODUCTION

The status of infrastructure in the United States has become a cause for concerns as the American Society of Civil Engineers (ASCE) has given the nation a letter grade of C- in their recent reports (*2021 Report Card for America's Infrastructure*, 2021). With the passing of the 2021 Infrastructure Investment and Jobs Act (IIJA), also known as the Bipartisan Infrastructure Law (BIL) through Congress, a surge of funding and investment into the nation's infrastructure has been implemented, with a focus to improve the quality and performance of the highway network across the United States. The bill calls for $1.2 trillion to be invested for the entire infrastructure and transportation system, with $550 billion of the total authorized for development of new programs and technologies (Rep. DeFazio, 2021).

State DOTs would benefit from this bill, as a major issue for these agencies is to analyze and evaluate the performance of the pavement networks in their jurisdictions in a timely manner. To achieve this, pavement engineers need to know more precisely which highway sections need proper attention to apply maintenance and rehabilitation (M&R) treatments, which would increase the pavement performance in the life cycle of these roads along with ensuring the safety of drivers vulnerable to the long-term effects of untreated distresses. To facilitate and accelerate the M&R decision making for which annual pavement data collection is conducted, many states have opted to use automated data collection of the pavement sections. Generally, a modern pavement management system (PMS) consists of the regular pavement data collection and the data-enabled M&R decision support. Currently, 33 state DOTs use automated data collection processes for their pavement networks for performance evaluation and strategic planning in revitalizing their transportation infrastructure (Luo *et al.*, 2022). However, the existing technology to accomplish has data quality issues with the accuracy of detecting and measuring the pavement distresses.

With the recent developments of artificial intelligence and machine learning (AI/ML), new techniques are being studied to confront the data quality issues of automated condition data collection for pavements. The most common methods using ML utilize image data to accomplish the task of pavement surface condition evaluation are object segmentation, detection, and classification. Proper training for models tasked with these objectives require large datasets of pavement images where ground truths must be established by engineers to ensure that a model

may accurately detect and classify distresses. The ability to differentiate separate types of distresses is essential for this effort to establish adequate performance of the models in real world conditions as the cause of distress may aide in pavement engineers determining what issues a highway section may be facing. Through extensive training, AI/ML models can accomplish this task when given sufficient data under a range of distress severity and the presence of multiple distress types, along with determining any non-distress objects found when collecting data.

The progression of these technologies is essential to the operation and safety of highway infrastructure systems. AI/ML have seen uses with the monitoring of structural health in buildings via vibration and signal waves (Rafiei and Adeli, 2018). However, a pavement network covers thousands of lane-miles, leaving the static and manual approach to be inefficient and inadequate for DOTs to execute. Through the vehicle-borne automated data collection methods, pavement engineers are able to cover a significant portion of the lane-miles. Such techniques can help gather valuable context for the condition and performance of a highway section that may require immediate maintenance and rehabilitation. The process for this strategic distribution of roadwork not only improves the quality and reduces life-cycle costs of the section, but it ensures the safety of the drivers as well (Eltahan, Daleiden and Simpson, 1999; Wang, Wang and Mastin, 2012; Li and Huang, 2014). In regions exposed to harsher climate and environmental conditions such as the oceanic or gulf coasts, the pavement sections are subject to saltwater corrosion, rainfall, extreme weather phenomena, and flooding at a higher frequency than roadways found inland and away from large bodies of water (Knott *et al.*, 2017). With the increasing intensity of these factors and events, the pavement structure is subject to loss of key integrity properties such as temperature resistance, aggregate adhesion, and deflection, leading to overall performance decline (Sultana *et al.*, 2018; Xiong *et al.*, 2019; Zhou *et al.*, 2021; Hong *et al.*, 2023). It is imperative that these sections receive adequate attention for the purposes of structural monitoring and condition evaluation, as these roads are vital for evacuation, supply chain, and commercial routes. With precise distress detection technologies available, pavement condition analyses and evaluations can be accomplished to provide context to engineers for strategic maintenance and rehabilitation operations.

In this research, deep learning AI/ML learning models are developed with an image library of pavement image data collected with a pavement survey vehicle. The image data consists of 2D

and 3D images in 4096x2048 resolution quality and are annotated with separate class lists based on surface type, containing distress items and other common objects on the pavement surface.

The objective of this research study is to develop modern, pre-trained DL object detection architectures on the collected 2D/3D pavement image data for efficient detection of pavement surface distresses and evaluation of pavement conditions for roadway sections in the Gulf regions. The research will aid in the effort of adoption of ML models by state DOTs and pavement engineers to receive accurate and reliable pavement surface condition data to support sound maintenance and rehabilitation (M&R) decision making to their highway pavements.

LITERATURE REVIEW

Pavement Distress Rating

The condition performance of a highway's pavement section can be evaluated through factors such as structural damage and roughness. For structural damage, surface distresses such as cracking and depressions are considered for their effects on the pavement's conditions. The geometry of each distress can indicate an issue with the section, whether it is from the material composition of the pavement layers, heavy traffic loads, age, resistance to environment, or water infiltration. A pavement rater's manual, such as the one written by the Texas DOT (TxDOT), describes how distresses typically appear.

Asphalt Pavement Surface Distresses

The most evident types of distresses that can be easily visualized on asphalt surfaces are transverse cracks, longitudinal cracks, block cracks, alligator cracks, and failures. The appearance of these cracks can be seen in Figure 1, with transverse cracks occurring perpendicular to the road's centerline and originating from declining temperature resistance, age, and/or loads, especially if the flexible layer was overlayed on top of concrete pavements with multiple joints. Longitudinal cracks are parallel to the centerline and extend down the roadway, which can be caused by tension failure in the surface layer along with loss of material strength from the base and subgrade layers. Block cracks are rectangular sections of the pavement that range from a 1x1-10x10 ft^2 area and can appear due to aging and extensive temperature cycles. Alligator cracks have tiny sections of the surface connected together in a web of cracks and are usually found along the wheel paths of

the road. Their origin is likely due to heavy traffic load, loss of base or subgrade materials due to water infiltration, or weak subsurface material strength. Failures appear as large voids on the pavement surface as large sections of the top layer are removed via erosion, widening of cracks, or depressions into the pavement. Failures such as potholes can be caused by heavy traffic loads, water infiltration into deeper layers, and/or unsupervised formations of small, localized cracks (*Pavement management information system, rater's manual fiscal year 2021*, 2020). Another common visible surface distress were sealed transverse and longitudinal cracks, where crack seal was applied as a form of M&R and are usually above the surface level by a few millimeters. Figure 1 lists these mentioned distresses from top to bottom, left to right, in the order listed.

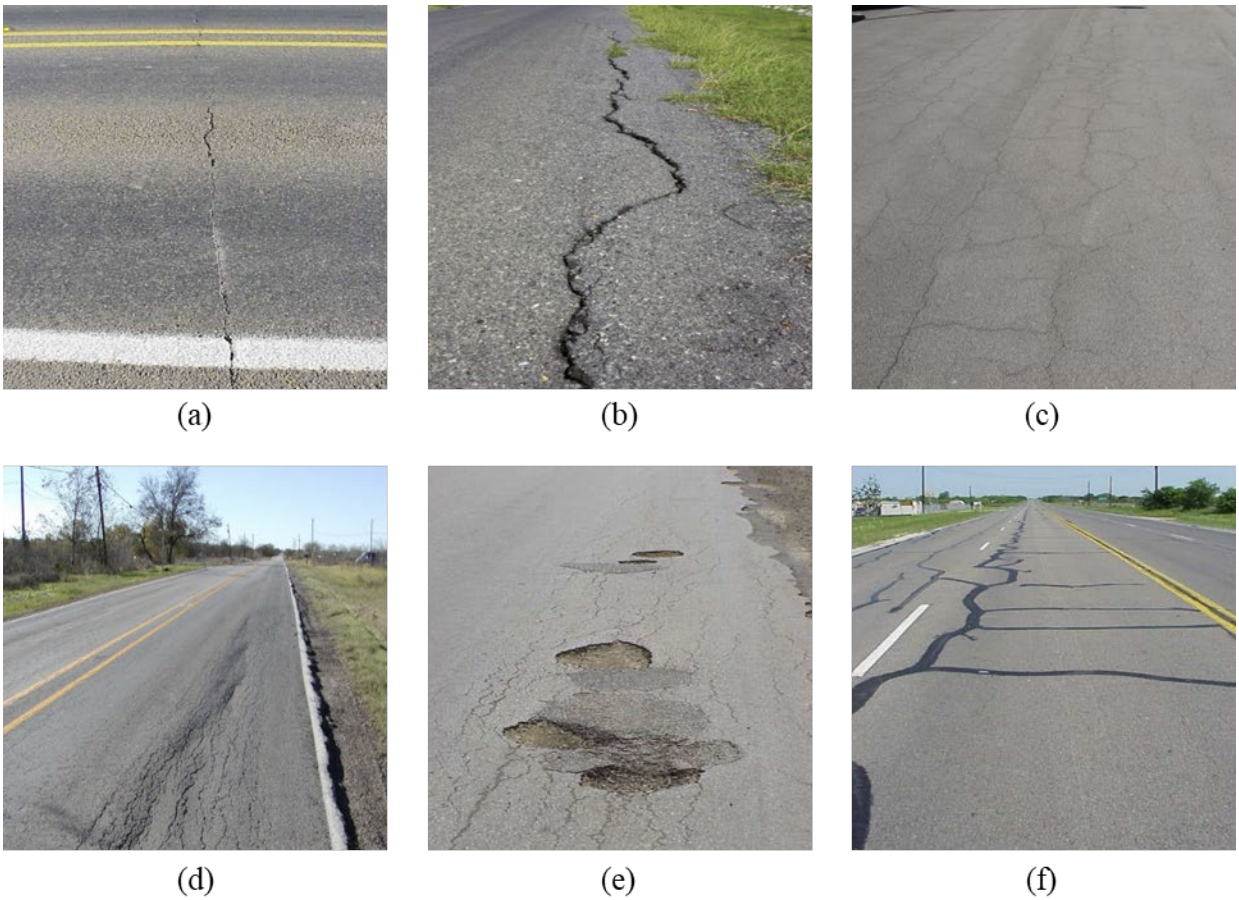


Figure 1. Distresses on flexible pavement: (a) Transverse crack; (b) Longitudinal crack; (c) Block cracking; (d) Alligator cracking; (e) Failures; (f) Sealed transverse and longitudinal cracks. Modified after TxDOT Pavement Rater's Manual 2021 (Pavement management information system, rater's manual fiscal year 2021, 2020)

These individual distress types are weighed differently when evaluating the condition of a pavement section. For instance, transverse cracks are formed with little to no traffic loading but

are not as severe as alligator cracks and potholes, which are formed due to heavy loading, water infiltration, loss of subsurface materials, or erosion.

Continuously Reinforced Concrete Pavement Surface Distresses

For sections of roadway that experience much heavier traffic loads and denser volumes of traffic, concrete pavement is used as an alternative to asphalt. Due to its much larger cost per unit and longer construction time, the material is used sparingly for sections that must yield higher performances for longer life cycles (*Pavement Manual*, 2021). When used for large highway distances, the most resilient form of concrete pavement is the continuously reinforced concrete pavement (CRCP). This type of pavement allows for concrete slabs several feet long with capabilities of spanning several miles before a singular transverse joint appears. The virtually uninterrupted length of a single lane slab is due to the continuous reinforcement from the steel rebar within the concrete, allowing for controlled crack spacing of transverse cracks along the surface and reducing the strain from transfer load across the joints.

Despite the physical strength of CRCP, constant, heavy traffic and intense loads due to its designed bearing capacity still create the circumstances for surface distresses to appear. Similarly to asphalt, transverse and longitudinal cracks do appear in CRCP. However, it is by design that transverse cracks appear much more often than longitudinal cracks as they are meant to appear in a controlled manner once the stress imposed on the pavement exceeds that of the concrete's strength. Longitudinal cracks would still appear, although rarely, due to improper traffic loads, poor steel placement, or inadequate subsurface layer strength (Hiller and Roesler, 2005).

Along with these two cracks that are found in both flexible and rigid pavement types, the most common distresses are spalled cracks, punchouts, asphalt patching, and concrete patching. Spalled cracks are formed when chipping occurs around a surface crack due to the infiltration of moisture into the concrete and steel, leading to rust and weakening the surface strength. Punchouts form when a longitudinal crack crosses two transverse crack to create a full depth block on the surface where faulting and spalling can occur. Since this full depth block causes uneven load transfer through the CRCP surface, shattering can occur in the punchout region and can lead to further spalling and faulting. Asphalt patches are temporary treatments to CRCP surfaces such as spalling and punchouts and can be installed as full depth repairs in localized spots around these distress regions. Concrete patches are a more permanent treatment for section areas with severe

distresses. These concrete patches are full depth and can occupy the entire lane width or more since they are cut cleanly into the slab. However, this causes the appearance of transverse joints and corners, leading to the appearance of joint failure and cracking localized within the patch (Choi *et al.*, 2020). Figure 2 displays the appearance of these distresses from left to right, top to bottom, along with failed joints (*Distress Identification Manual for The Long-Term Pavement Performance Program (Fifth Revised Edition)*, 2014; *Pavement management information system, rater's manual fiscal year 2021*, 2020).

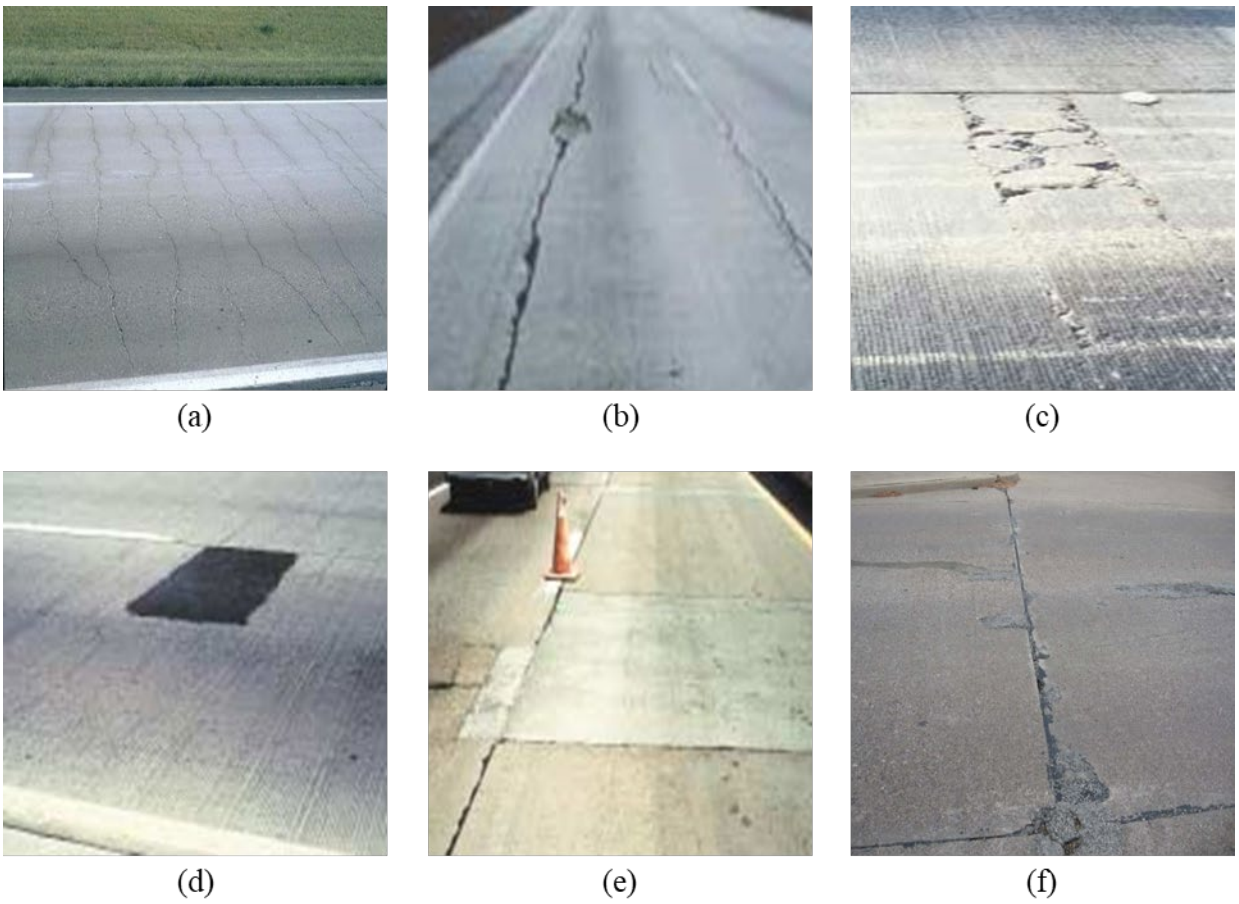


Figure 2. Distresses on rigid pavements: (a) Transverse cracks; (b) Longitudinal cracks; (c) Punchouts with severe chipping and spalling; (d) Asphalt patch; (e) Concrete patch; (f) Joint failure. Modified after FHWA Distress Identification Manual for the Long-Term Pavement Performance Program, Fifth Edition (Distress Identification Manual for The Long-Term Pavement Performance Program (Fifth Revised Edition), 2014) and TxDOT Pavement Rater's Manual 2021 (Pavement management information system, rater's manual fiscal year 2021, 2020)

Although uncommon and not generally considered as a distress in CRCP, they can still occur either due to transverse joints from construction limitations, concrete patch placement, or load transfer from vehicles switching lanes.

For concrete pavements, transverse joints and smaller slabs are more common in jointed concrete pavements (JCP). While data was collected for JCP sections, they were not included in this study due to availability of the annotated images of the surface type. CRCP sections were given priority as the rigid pavement surface for the study as it occupies more lane-miles in the highway network due to their resilience compared to JCP (*Pavement Manual*, 2021).

The quantity and severity of the distress affect the condition score of a section, which engineers and DOTs can use to determine when to apply M&R. The use of M&R has to be strategic, at intervals during a pavement section's life that may extend its life cycle while being cost-effective and efficient for the roadway's performance. This approach for improving the quality of the highway infrastructure can provide much needed repair while minimizing the resources needed to perform such actions when conditions become too severe to prove effective. However, the rate of data collection for PMS databases across several states requires fast and accurate methods to properly evaluate the pavement surface within a reasonable time.

Pavement Data Collection

There are three ways to collect pavement surface condition data, which include manual, semi-automated, and fully automated processes. Manual data collection involves the use of measurement tools and sample sections of roadways to evaluate the conditions of the pavement surface. Semi-automated reduces the need for human intervention during the data collection process as the entire section's surface can be collected with visual data such as images but still requires measurements taken from engineers once section data is collected. Automated methods to collect data are used in 33 state DOTs in order to cover as much ground as possible for condition evaluation, but the process still requires human verification as the technology still faces issues with detection accuracy and distress measurement.

The push from manual and semi-automatic to automatic pavement surveying can be found during the 1990's, as the traditional methods had caused issues on the roadways. The manual data collection was dangerous for the personnel collecting the data and yielded a high-cost rate for the

operations. In 1997, Cheng and Miyojim worked on techniques to provide an image enhancement algorithm along with tasks to detect and classify distresses. Multiple experiments were run with their algorithm and had returned satisfactory results for what was a novel approach to the issue, heavily relying on a custom skeleton configuration analysis (Cheng and Miyojim, 1998). In 2008, Tighe et. al had collaborated with the Ministry of Transportation Ontario and three companies in Ontario, Canada to determine the effectiveness of automatic and semi-automatic data collection techniques compared to manually collected data. They had utilized laser, infrared and ultrasonic beam sensors along with accelerometers, distance measuring systems, and digital imaging to capture pavement surface data and evaluate the condition of the road surface using values from their distress manifestation index (Tighe, Ningyuan and Kazmierowski, 2008). Tighe et. al had found that the automated systems they used were comparable to the manual evaluations at the time, as well as being more compatible to flexible pavements than rigid and treated surfaces. Due to the early progress into the field of automatic data collection and evaluation covered by this paper, the author had stated that the automated surveys be supplemented with the manual surveys to establish more accurate pavement condition analyses.

An issue with the automated data collection process for pavement sections is the data quality and equipment used during both the initial collection and evaluation of the surface distresses. Using an online questionnaire survey, Luo et al. investigated the techniques and methodologies that were conducted by several local and state highway agencies. It was found that there was little to no uniform protocol for the automated or semi-automated data collection and that data quality proved to be a major issue to the agencies. Along with this, novel technologies such as AI would need to be improved and specialized for pavement object and distress detection before full-scale implementation across pavement networks (Luo *et al.*, 2022).

Convolutional Neural Networks

Most, if not all, data collected from the pavement surface from automated processes should be in the form of images. Image data can be processed through convolutions as the pixels and their intensity values serve as a matrix that would train a model utilizing this technique. Neural networks are adept in processing visual data as the neurons in the network behave in a similar manner to the human brain for contextualizing data (Smola and Vishwanathan, 2010). In a neural network, the input image pixels behave as a matrix, with each cell's value holding the intensity of that pixel.

Feature maps are created during the convolution process as the kernel combs through the image matrix. As shown in Equation 1, the kernel's dimensions dictate its starting point in the top left corner of the image where the process of convolution creates products of the elements within before summing the values to create a new cell value. Once that region of the image has been altered, the kernel may move to the neighboring cell, depending on the value of the stride given for the kernel to travel through the matrix.

$$(I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) \cdot K(m,n) \tag{1}$$

I(i,j) is the input matrix, I, of the image, with i and j representing the row and column positions of a pixel. K(m,n) behaves similarly, with m and n being the kernel's, K, coordinates as it combs through the image. The summations are the iterations the kernel goes through as it convolutes the values of the matrix.

Once the kernel completes a given number of convolutions in a layer, an activation function, such as the rectified linear unit (ReLU) in Equation 2, is used to introduce non-linearity within the feature map, eliminating any negative values that may be produced. Then, pooling, or downsampling, of the feature map in Equation 3 is executed to reduce the spatial dimensions of the input image to save on computational power and aid the algorithm determine variations to certain classes based on this variation (Smola and Vishwanathan, 2010; Prince, 2023). This process is repeated to a certain point and is reversed using Equation 1 once the feature maps are transferred through each layer. When the output feature maps are brought together in a fully connected layer, the activation function in Equation 4 links each output layer through a softmax sigmoid activation where class probability is determined and distributed.

$$ReLU = \max(0, x) \tag{2}$$

$$O(i,j) = \frac{1}{n * m} \sum_{p=0}^{n-1} \sum_{q=0}^{m-1} I(i+p, j+q) \tag{3}$$

$$\sigma(\underset{z}{\rightarrow}_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_i}} \tag{4}$$

For Equation 3, the output, O, at a given coordinate of the image (i,j) can be formulated as the input values at I(i+p,j+q) are subject to the pooling operation based on the dimensions of the

pooling layer, m and n, where p and q are coordinates that work on the pooling layer. The activation through the softmax function σ in Equation 4 is determined by the input vector $\underset{z}{\rightarrow}_i$ where the input itself is calculated for its probability to appear in a K set of classes following the output z_j (Smola and Vishwanathan, 2010). With these functions and operations combined, CNNs are adept at handling image data and can perform well with their tasks in object detection and item classification. The efficiency of a CNN is dependent on the image quality along with the ability for the network to discern key features to effectively detect any noticeable items within the dataset.

Transformers

One common machine learning framework is a transformer, a self-attention-based architecture. While it originated in use for natural language processing tasks, it was used with CNNs to improve accuracy. As the architecture was further developed, the capabilities of transformers conducting image scaling processes on their own became more feasible, with Dosovitskiy et al. developing a Vision Transformer (ViT) to adequately compete with other state-of-the-art models. They altered image processing sequence from self-attention and inductive biases to perceiving the input images as sequences of patches. This led to their team finding relative success with image classification with substantially less computational costs (Dosovitskiy *et al.*, 2021).

Other advancements with transformers went beyond image classification and towards more complex tasks such as detection. Carion et al. at Meta AI developed a new framework for image transformers using single pass predictions of object and set prediction loss which would force unique matches between the ground truth and predicted input data (Carion *et al.*, 2020). The architecture for their transformer, Detection Transformer (DETR), contained a CNN backbone for feature extraction, an encoder-decoder transformer, and a feed forward network to make final predictions. Along with detection, DETR was capable of panoptic segmentation and had comparable results to Faster R-CNN at the time of study.

There was an issue with DETR where it traded accuracy for computational costs and speed. This was due to the exclusion of non-max supression (NMS), which would select the most relevant predictions to improve speed while not optimizing for accuracy. NMS is commonly used in You Only Look Once (YOLO) models, which are known for their balance of speed and accuracy. Zhao

Y. et al. (2024) investigate this discrepancy between DETR and YOLO to develop a model that could equal YOLO in terms of speed for real-time object detection (Zhao *et al.*, 2024). Their model was called Real-Time Detection Transformer (RT-DETR) and expanded on the feats of the original DETR by creating a hybrid encoder for multi-scale feature processing to improve speed. To maintain accuracy, the decoder is given high-quality initial queries from the uncertainty-minimal query selection after feature fusion in the encoder. They had found that RT-DETR was able maintain sufficient accuracy while allowing for flexible training speeds comparable to the YOLO line of models. RT-DETR proved useful with its improved speed and accuracy when He et al. used the model as part of its study with other state-of-the-art architectures to detect diabetic retinopathy in patients. The challenge proposed from being able to accurately detect the damaged blood vessels in the eyes from the other small, densely-packed targets (He *et al.*, 2025). The results from the study showed the capabilities RT-DETR had with complex visuals with multiple non-background objects. This feat establishes the viability of the transformer model to perform well with minute details surrounded by numerous objects, which is useful to the study of this research.

With the advancements and complexity of DL model architectures, applications of multi-modal datasets for AI/ML development seem more feasible. The ability to process image channels can provide access for deeper feature extraction and smoother object detection. Using multi-modal data within a single input provides valuable context when training a DL model when attempting to circumvent issues with an image layer.

2D/3D Image Utilization

Visual data is subject to data quality issues such as corruption, distortion, noise, and lack of distinction for material characteristics. This is evident in 2D images of pavement surfaces where oil stains and debris tend to cause confusion to AI/ML models to incorrectly label the objects as distresses. With proper techniques, the distinctions between similar objects and materials can be circumvented to achieve promising results. A review conducted by Voulodimos et al. went through multiple significant approaches to computer vision in deep learning. They noted that CNNs often require supervised, labelled data to perform at an optimal rate to withhold its ability of feature learning. Along with this, other types of model architecture could perform unsupervised learning or had less computational costs, but did not have the capabilities CNNs had to be less susceptible to errors when input data transformation was applied (Voulodimos *et al.*, 2018). Qi et al. utilized

a novel visual blurriness network for the detection of glass surfaces in 2D images by focusing on the slight distortion caused by the refraction of light (Qi *et al.*, 2024). Aside from visible light, networks can be developed to extract feature data from multiple channels. Tang et al. used modalities which would process RGB and thermal visual data separately in their model. Once both processes were complete, both modalities were integrated for feature extraction, significantly improving detection performance (Tang *et al.*, 2025). Liu X. et al. used Domain Adaptive Object Detection (DAOD) to detect missed annotations due to instance ambiguity such as the presence of noise. With the use of a Noise Latent Transferability Exploration framework, the accuracy of the DAOD model with the dataset used in their study significantly improved (Liu *et al.*, 2022).

Many of the major developments in the field of AI/ML for pavement distress detection have occurred within the past 15 years. Ouyang and Xu worked on an automated pavement detection system in 2013 that utilized images taken from a 3D camera coupled with a laser projector that captured transverse profiles of the road. The equipment had a pixel depth of 0.5mm at a height of 1.4m above the road. The longitudinal resolution of their data was dependent on the speed of the vehicle during operation (Ouyang and Xu, 2013). Ouyang and Xu opted to evaluate the accuracy of their automated system on a single section to determine the repeatability of the 3D capture equipment, where they found a 2% difference in crack length. While the reliability of 3D image data had been shown to help algorithms detect distresses more optimally, Huang et. al opted to combine the characteristics of both 2D and 3D for their model. Using the Dempster-Shafer Theory to address any uncertain or missing data in the dataset, two datasets were used, one with 220 images and 104 cracks where the other set had 155 images with 42 cracks (Huang, Liu and Sun, 2014). Their approach was to train their models using only 2D data, 3D data, and a fusion of the two. The findings of the experiments had shown that the 2D only approach yielded higher error rates in the first data set, with the 3D only approach establishing uncertainties in both datasets and higher errors rates in the second data set. The 2D-3D fusion of data in the model had returned higher accuracy and more precise detections along with the elimination of uncertainties across both datasets. Liang et al. stated the issues with using pavement photographs as they were susceptible to lighting and errors from shadow interference. They used photographs and infrared sensors along with multisource image fusion based on similarity and difference (MSFSD) to combine the features of both visible light and infrared for better crack detection (Liang *et al.*,

2024). Using this image fusion method provided much higher accuracy and detection efficiency than other prosed methods.

ML Use For Pavement Distress Detection

Massive strides in the utilization of deep learning for pavement distress detection when Zhang L. et. al used ConvNet as their model to accomplish this task. Their model was trained using 500 images taken from a smartphone at a resolution of 3264x2448 pixels. After conducting sampling with 99x99 RGB image patches for each image, their dataset was increased to using 640,000 samples for training, 164,000 for validation, and 200,000 for testing their model. The findings reveal that the ConvNet model performed well, with a precision score of 87%, 92.5% for recall, and 89.71% for F1 (Zhang *et al.*, 2016). A year later, two more experiments were made by different groups for pavement crack detection using 3D imaging. Zhang A. et. al used CrackNet as their detection model, an AI/ML architecture based on CNNs that differed from others as it does not use pooling layers for feature extraction. Rather, CrackNet uses predicted class scores for all pixels that it detects.

Zhang A. et al trained their model with 2,000 randomly selected 3D pavement images from a library with a 9:1 ratio for training and testing. CrackNet in this paper was compared to Pixel-Support Vector Matrices and 3D shadow modeling and had performed much higher with its precision, recall and F1 scores, standing at 90.13%, 87.63%, and 88.86%, respectively (Zhang *et al.*, 2017). However, one of the main issues with CrackNet was its difficulty in processing thin hairline cracks. The other paper released in 2017 was by Tong et. al, where the combination of ground penetrating radar (GPR) and CNNs were used for the detection and measurement of concealed cracks below the pavement surface. They had 3 different models to perform dedicated tasks such as recognition, location, and measurement of the concealed cracks under the pavement surface. Around 500 256x256 grayscale maps and 6832 images, which included distresses, concealed cracks, subgrade settlement, and roadbed cavities were used as training samples for the CNNs (Tong, Gao and Zhang, 2017). Tong et. al had found that the recognition and location CNNs had performed well in their tasks with little to no error, but the measurement algorithm could not provide precise 3D reconstructions of the cracks as the feature extraction methods did not meet the desired results.

Another approach to distress detection would be through its use in pixel segmentation of pavement surface images. Akagic et. al collected 50 images through an Internet search, containing wide variation of factors such as lighting, angle, distance from surface, distress severity, and image quality. Using Otsu thresholding, the foreground and background data had become separated within the image along with the images being divided into four equally sized subimages. After this, the RGB pixel locations were recorded on a local and global level given their position on the original image. Then the images were converted to grayscale for detection in wet and dry conditions of the pavement surface. Akagic et. al stated that their technique provided rough estimates for visual evaluations of the roadway surface. Due to the wide range of factors that affected the image dataset, the findings by Akagic et. al served more as a showcase of crack pixel segmentation rather than effective performance for in situ conditions (Akagic *et al.*, 2018).

As advancements were made in the field of object detection for AI/ML models, Gong et al. investigated the factors within a dataset which could play a major role in determining a model's performance. Comparing a YOLOv5 model with Faster R-CNN, the F1 scores were raised from 0.7 to 0.84. This improvement led Gong et al. to conclude that due to the unique nature of distress detection as a form of object detection with less distinct features amongst classes required much more effort to achieve adequate scores, urging for more sufficient dataset sizes and consistent annotation work (Gong *et al.*, 2023).

You Only Look Once (YOLO) Models

Liu J. et. al created a complex algorithm utilizing YOLOv3 and a modified version of U-Net to accomplish this fusion of detection and segmentation called a two-stage approach. For detection, the model used was YOLOv3 with a Darknet-53 backbone while the segmentation half of the architecture consisted of a modified U-Net backbone and a pretrained ResNet-34 as the encoder. The images for the dataset were taken at a height of 1.5 meters above the pavement surface, 1,066 images were collected around the city of Changsha in the Hunan Province of China (Liu *et al.*, 2020). A secondary set of images were extracted from the CrackForest Dataset (CFD), which consisted of 118 images of 480 pixels by 320 pixels taken by a cellphone in Beijing, China. After several augmentations, the total dataset resulted in a collection of 27,966 images. Comparing multiple papers along with a one-stage approach, it was found that the two-stage approach yielded much higher precision and F1 scores than the other models. The recall score was only lower by

0.0139 than the model who held the higher value. Another use of YOLOv3 was when Du et. al collected over 45,000 images of flexible pavement surface, with a total of 59,366 pavement distresses present across 200 km of road surface while attempting to avoid heavy traffic congestion. These 1920x1080 pixel images were taken at a speed no faster than 80 km/h, or 50 mph and were labelled using the labelImg-master_v1.5.2 for classification on cracks, potholes, net cracks, patched cracks, patched potholes, and patched net cracks. Du et. al had compared their YOLOv3 model by using a batch size of 64, batch size of 96, a Faster R-CNN algorithm, and a Single Shot MultiBox Detector (SSD). The experiment yielded that the YOLOv3 model was efficient at detecting manholes, patched cracks, and patched potholes, performed better than the SSD model, and was similar in precision and F1 to the Faster R-CNN. The models, however, had difficulty detecting and classifying cracks, potholes, net cracks, and patched net cracking with scores below 0.60 (Du *et al.*, 2021).

With the capabilities of YOLOv3 for pavement distress detection taken into consideration, newer versions of the CNN were also evaluated for the performance of detecting surface distresses. The major improvement in detection capabilities from YOLOv3 to YOLOv5 in general were quite significant. Karthi et al. set out to examine the strength of YOLOv5 using five datasets containing images of either fish inside aquariums, sign language letters, chess board pieces, raccoons, and books. The consistency of these datasets determined YOLOv5's capabilities for wide ranges of dataset sizes, quality, input dimensions, and class distribution (Karthi *et al.*, 2021). Their findings claimed that YOLOv5 is a robust model fit for various feature extractions and precise classifications. With that taken into account, the model can aid in the effort for accurate pavement distress detections. Hu et. al wanted to see the viability of YOLOv5 for this task and assembled four versions of the model, varying in size from small to extra-large. The paper states that 3001 images of 2976x3968 pixel quality were divided into 1920 images for training. 480 for verification, and 601 for testing. Detection was the sole purpose behind the study of these YOLOv5 models so only a single distress class was used (Hu *et al.*, 2021). The backbone of these algorithms used a cross stage partial network (CSPNet) for the backbone, and path aggregation network and feature pyramid network were installed for the neck. It was determined that YOLOv5 was highly efficient for the task of pavement distress detection, with the smallest model being the fastest and the large model having the highest precision scores. After this performance experiment, Ren J. et. al included the task of classification of longitudinal cracks, transverse cracks, alligator cracks, and

potholes. Images for the dataset were taken using a dashcam with a resolution of 1920x1090 pixels at a frequency of at least 10 Hz. The CNN was modified with a revised CSPNet as the backbone to reduce the computational load of training and generate cross channel feature maps, and four different attention modules that were evaluated as well. These included Convolutional Block Attention Module (CBAM), Coordinate Attention (CoordAtt), Squeeze-and-Excitation Network (SENet), and Efficient Channel Attention Network (ECANet) (Ren *et al.*, 2022). After training with 500 epochs, a batch size of 64, a learning rate of 0.01, and 7-hour sessions, Ren J. et. al found that the attention modules had adequately improved the metric scores of YOLOv5 outside of recall, with the CoordAtt module yielding the highest scores. By developing a Central Feature Pyramid (CFP) network, Quan et al. modified the YOLOv5 and YOLOX models to improve their performances on the MS COCO 2017 dataset. The scores of the two had improved due to the CFP network's capabilities in applying focus to the corner regions and deep features of the input images, regulating the shallow, central features within (Quan *et al.*, 2023).

As development of newer YOLO models came into fruition, their object detection capabilities would advance as well. With the release of YOLOv8, Varghese & Sambath evaluated the model through rigorous benchmarks compared to other contemporary state-of-the-art models with COCO, PASCAL, VOC, and WIDER FACE datasets. Through their studies, it was found that it had greatly surpassed previous YOLO models in terms of precision and efficiency, with plans to implement the technology on other devices (Varghese and M., 2024). A study by Yi et al. would incorporate YOLOv8 into their own model, LAR-YOLOv8, which included alterations to the feature extraction network, feature pyramid network, and loss function. Using three open source datasets (NWPU VHR-10, RSOD, CARPK) which consisted of aerial or satellite photos of random locations, aircraft, or parking lots, their team had found that LAR-YOLOv8 found success in accurate detections of small objects in remote sensing images contained within those datasets (Yi *et al.*, 2024). Additionally, findings from Wang H. et. al have shown the advancements of the YOLO CNN model with YOLOv8. A dataset of 2000 sheets was used to train the model with an 8:2 ratio for random sampling. Along with a standard YOLOv8, the model was compared with a regular YOLOv5 architecture and modified YOLOv8 that includes a Darknet53 backbone with a CBAM at the neck. This altered model was named YOLOv8-D-CBAM and had more modifications such as backbone layers replaced with depth-separable DWConv with the neck adding channel attention mechanisms (H. Wang *et al.*, 2024). The input size of the data was

320x320 pixels with differing output sizes of 10x10, 20x20, and 40x40. The models were tasked with detecting and classifying transverse, longitudinal, blocky, and irregular cracks on the pavement surface. Standard YOLOv8 metrics performed better than the YOLOv5 model's while the YOLOv8-D-CBAM model excelled in each category of precision, recall, F1, and mAP with scores ranging from 0.982-0.995.

Within the last year, three newer versions of the YOLO line of models have been released. These are YOLOv9, YOLOv10, and YOLOv11. YOLOv9 had massive changes to its architecture, utilizing a General Efficient Layer Aggregation Network (GELAN). This addition to the model's architecture provided enhanced detection capabilities, with Xu et al. developing a GELAN algorithm to monitor outbreaks of violence in crowded sceneries and Li J. et al. using YOLOv9 to develop a model for mitigating the effects of adverse weather conditions in detection of power lines (Li *et al.*, 2024; Xu *et al.*, 2024). With the task of pavement distress detection, Youwai et al. utilizes a dataset stemming from multiple countries where road surface distresses includes transverse cracks, longitudinal cracks, alligator cracks, and potholes to evaluate YOLOv9tr in comparison to other models such as YOLOv8, YOLOv9, and YOLOv10. Achieving a frame rate of up to 136 frames per second, YOLOv9tr proved adequate for use in automated pavement distress detection and scored higher than most of the other models in the study with fewer parameters (Youwai, Chaiyaphat and Chaipetch, 2024).

YOLOv10 was developed to offer higher training speeds than YOLOv9 by offering an NMS-free approach in its design. Qiu et al. developed an altered version of YOLOv10, LD-YOLOv10, for object detection from drone-captured data, where target size is small and object density is high. The modifications to the backbone and neck via RGELAN and DR-PAN structures, respectively, allowed for more refined feature extraction while improving efficiency in the model, leading to slightly higher accuracy scores (Qiu *et al.*, 2024). With its similarity of capabilities in scope for pavement distress detection, Sohaib et al. evaluated several models for the presence of cracking in concrete structures, including YOLOv10. Through transfer learning, crack detection and segmentation dataset are used and provide insight to the effectiveness of YOLOv10, as it had outperformed the previous versions of the YOLO line with YOLOv10x (Sohaib, Arif and Kim, 2024).

When YOLOv11 was released, it opted for more refined accuracy than that of YOLOv10, where training and inference speed was focused on. With intention to bypass interference and low efficiency in underwater visual data, Cheng et al. developed and improved YOLOv11 for underwater target detection. Through image preprocessing via adaptive lighting and actuators, the YOLOv11 model that was developed was able to slightly improve accuracy with fewer parameters on underwater targeting datasets (Cheng *et al.*, 2025). With this low visibility study on YOLOv11 taken into consideration, further use of such applications were conducted using YOLOv11. Zanevych et al. proposed a modified version of YOLOv11 using feature pyramid networks and Grad-CAM++ called YOLOv11+FPN+Grad-CAM for detection of potholes in low visibility conditions such as evening or nighttime. Compared with models such as YOLOv8, YOLOv9, YOLOv10, RT-DETR and MobileNetSSD, it was found the modified YOLOv11 was able to outperform the other models in the mAP50-95 score metric (Zanevych *et al.*, 2024).

Other Architectures

Apart from the use of YOLO, other methods such as image preprocessing have been used to optimize the precision of the detection algorithms for pavement distresses. By using a modified Faster R-CNN with a Residual Network (ResNet) and its RPNs for feature extraction, Zhao M. et. al sparsed their images with denoising, grayscale, thresholding and contrast enhancement techniques. The sparsing of the images were paired with algorithms that accomplished one of the techniques and were compared with the regular images from the dataset of 150 total images, with 90% dedicated to training. The experiments were held for the model using regular images, models paired with one other algorithm, paired with two algorithms, or all algorithms being applied. The results of the paper show that the use of some algorithms to sparse the images had improved the metrics scores by at least 5%, despite the presence of noise and degradation of quality within the images (Zhao *et al.*, 2022). Another approach to pixel-level detection was made when Guan et. al opted for the use of Structure from Motion (SfM) to collect pavement data for their modified U-Net model. The SfM technology enabled the use of 3D recreations of the roads surface using 2D images taken at different locations of the scene. Three GoPro Hero 8 cameras were used to accomplish this task when Guan et. al collected data on the roads of Chang'an University and Xi'an in the Shaanxi Province of China (Guan *et al.*, 2021). The acquired dataset yielded a total of 600 512x512 pixel images, divided into three equally sized subsets for color, depth, and

overlapping images. When evaluated, their U-Net model had reached scores around 0.85 for precision, recall, and F1 metrics. Other models such as U-Net have also seen modifications geared towards pavement distress detection and segmentation tasks. The developments of CrackNet led Huyan et. al to test the capabilities of ResNet and Visual Geometry Group Network (VGGNet) when applied to the neural network. Around 500 images were collected via smartphone and action cameras held 1.3m above the top layer along with a pavement condition survey vehicle with the equipment being 2m above the road surface. These images were randomly sorted into a ratio of 3:1:1 for training, validation, and testing of the updated CrackNet models (Huyan *et al.*, 2022). Two classes were chosen for the list of distresses in the dataset; linear cracks consisting of transverse and longitudinal cracks and map cracks, which included fatigue and block cracks. To further evaluate the effectiveness of the two models, they were compared with a fully convolutional network, a pyramid scene parsing network (PSPNet), a base U-Net, and a standard CrackU-Net. Huyan et. al had determined that both VGGCrackU-Net and ResCrackU-Net had outperformed the other neural networks for using pixel-level crack detection. Between the two prototype models, ResCrackU-Net had higher F1 scores than VGGCrackU-Net by 0.01-0.02.

Transformers have also provided the capacity to effectively detect pavement surface distresses. Lin Z. et al. developed a network based on a visual transformer and self-supervised learning to accomplish the task on pavement anomaly detection in a dataset of 4,428 images. The dataset contained an even split between normal images and images with anomalies present within. The team had found the approach of self-supervised learning with the vision transformer highly accurate with results comparable to data augmentation (Lin, Wang and Li, 2022). Guo X. et al utilized DETR along with Faster R-CNN, YOLOv5, and YOLOv8 to detect cracking and looseness using images gathered via radar waves from GPR on a single lane track road. This approach was to determine the ability of the models' performances in nondestructive testing of subsurface distresses. Despite having the worst performance of the four for both precision and efficiency, the study was an early venture into the use of DETR for civil engineering tasks, whereas the YOLO models performed much more adequately (Guo and Wang, 2024). With the modifications to DETR presented by RT-DETR for real-time detection speeds, Liu G. et al. improved the RT-DETR model for surface crack detection. Their methods included alterations to the backbone and multi-scale feature fusion modules to optimize crack feature extraction and minimizing the presence of noise

within the input data from the China-D dataset. Their results allowed for better focus on crack regions with a slight improvement in mAP50 scores by 1% (Liu *et al.*, 2024).

METHODOLOGY

Data Library

Asphalt Pavement Data Collection Sites

With the image data utilized throughout the literature review, a custom data library was needed for this research to develop an AI/ML model. The pavement surface data collection was a multi-state process, with roadway sections collected in Texas, Louisiana, and Mississippi. Approximately 408 miles of asphalt pavement data were gathered in the coastal plains of the neighboring states of Louisiana and Mississippi. The locations of the sections in these states are not solely for proximity to the coast, but their susceptibility to inundation provided an opportunity to collect data on pavements with reduced service life from severe weather (Elseifi, Mousa and Gaspard, 2022). As for the sections in Texas, nearly 31 miles of asphalt pavement were collected inland in the counties of Hays, Travis, and Guadalupe, near the I-35 corridor. The inland sections provide an opportunity to analyze and compare the performance of the pavements near the coast and in the coastal plains with roadways in drier conditions with less exposure to the coastal environment.

The sections where the data was collected are shown in Tables 1 and 2 as well as Figure 3. Although some sections in Louisiana and Mississippi are more than 100 miles from the coast, they are still located within the coastal plains for the Gulf of Mexico while the sections that are inland in Texas are on the edge of the geographical boundaries, per the findings of W. J. Sutherland (Sutherland, 1908). Further data selection was conducted to remove images with high noise or extremely dense presence of distresses with varying severity. To collect the surface images of these sections, the pavement survey vehicle with a HyMIT sensor as shown in Figure 4 was used.

Table 1: Louisiana and Mississippi coastal pavement sections

Section	Location	Length (miles)	Images	Defining Distresses
LA-27	Sulphur/Hackberry, LA (Calcasieu/Cameron Parish)	12	513	Block, Alligator
LA-14	Lake Arthur/Kaplan, LA (Jefferson/Vermillion Parish)	26	612	Block, Failures
US-90	White Kitchen/New Orleans, LA	10	567	Longitudinal, Block
MS-53	Gulfport/Perkinston, MS (Harrison/Hancock County)	20	593	Longitudinal, Block
MS-13	Lumberton, MS	10	213	Transverse, Longitudinal
Holly Bush Road	Brandon/Pelahatchie, MS (Rankin County)	6.5	238	Block, Alligator

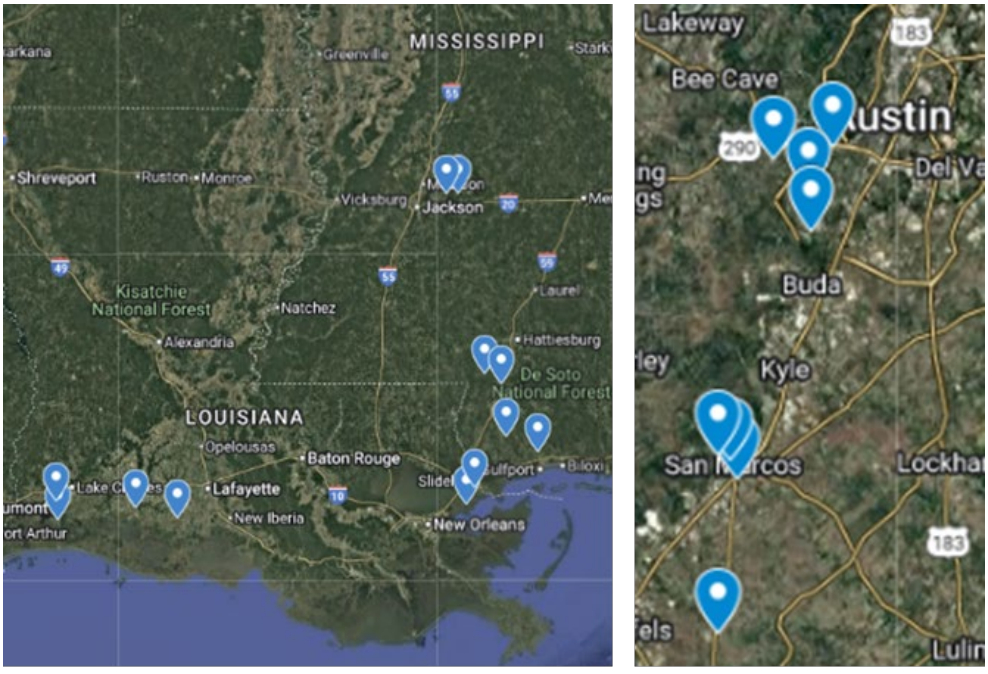
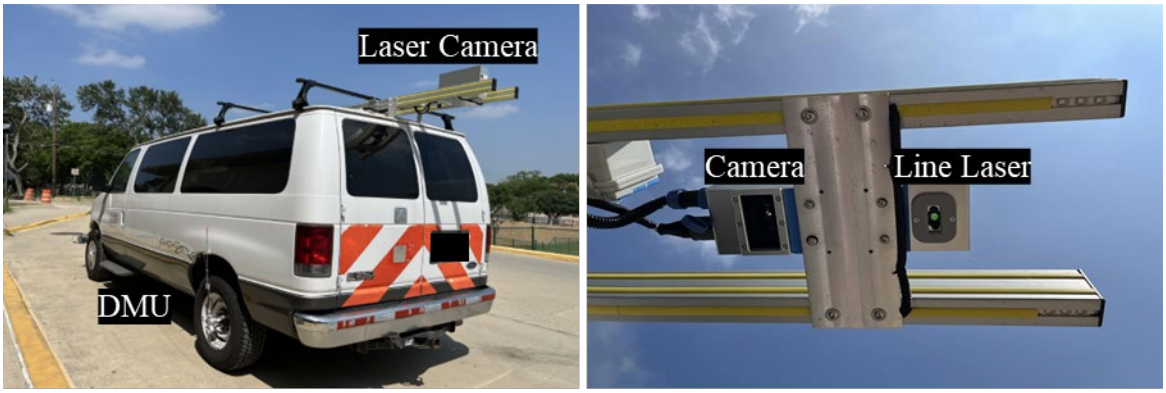


Figure 3. Location of asphalt pavement data collection sites: (a) Asphalt sections in Louisianan and Mississippi; (b) Asphalt sections in Texas.

Table 2: Texas inland asphalt pavement sections

Section Name	**Location**	**Length (miles)**	**Images**	**Defining Distresses**
Old Ranch Road-12	San Marcos, TX	2.6	306	Transverse, Longitudinal, Alligator
TX-123	Hays/Guadalupe County, TX	7.8	881	Transverse, Sealed, Longitudinal,
West Slaughter Lane	Austin, TX	16.2	1818	Block, Alligator
Brodie Lane	Austin, TX	3.5	387	Failures, Block Alligator



(a) (b)

Figure 4. Pavement survey vehicle: (a) Position of distance measurement unit (DMU) and line laser sensor camera; (b) Line laser and camera.

Attached to the vehicle was a HyMIT HY3D4K laser sensor, which would capture visual data of the road surface via 2D images and their respective 3D laser scans as the driver operated the vehicle. A strong line laser was emitted from the sensor spanning 14’ across the transverse direction at a longitudinal resolution of 7mm and 1.042mm in the transverse direction. With this, a single image of the pavement surface covers 47’ longitudinally and the pixels where distresses were present would return a darker intensity from the rest of the pavement surface while objects

with higher elevation from the road appeared brighter in 3D images. The data was recorded perpendicularly to the direction at which the vehicle was travelling, which minimized the error for distress profiles to be obscured from how the sensor was angled with respect to the surface of these sections. Combined, approximately 45,000 images of pavement surface data were collected across the three states.

The library of image data included the distresses shown in Figure 1, as these were the most common forms of cracks and potholes present in flexible pavement surfaces. Along with collecting a substantial number of distress data with the 2D and 3D images, other common objects in the dataset had to be taken into consideration as displayed in Figure 5. The most prevalent of these were transverse joints in the asphalt and lane longitudinal cracks, or longitudinal joints, which are a result of lane construction when the pavement was created.

(a) (b)

Figure 5. Non-distress objects on flexible pavement surfaces: (a) asphalt joints; (b) lane longitudinal cracking.

By distinguishing these distress and non-distress objects, detection algorithms have the capability to reduce the occurrence of false positives when attempting to predict the location and classification of real distresses. Other issues may arise during the data collection, such as the appearance of noise in Figure 6. This drop in image quality can reflect the conditions of data collection in the field, where changes in lighting from weather and noise affect the accuracy of

real-time detection. While these reduce the quality of the dataset, their inclusion can help a model be more suitable in applications for automated data collection where there are limits to image preprocessing due to loss of accuracy.

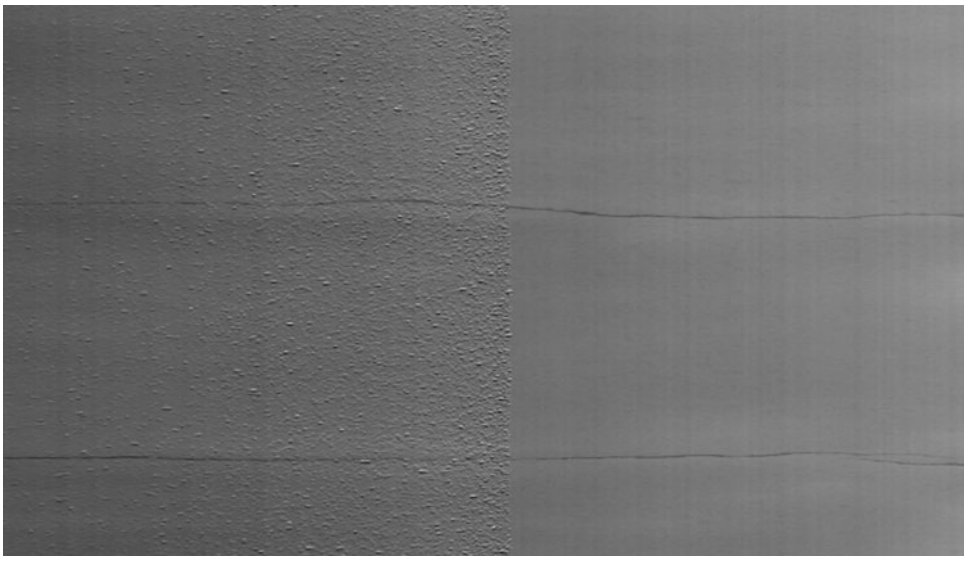


Figure 6. Noise present within an asphalt surface image.

Continuously Reinforced Concrete Pavement Data Collection

Although consisting of significantly less lane-miles than asphalt, CRCP roadways are critical sections of the highway network in coastal regions. As they are constructed to perform under much heavier environmental and traffic loads for longer service periods, CRCP is used in major urban areas, energy sector corridors, toll roads, and bridges. These vital functions limit the expensive use of this pavement type due to the construction time and material costs, along with the serviceability of the highway.

One long section of CRCP was collected south of Houston, TX in Brazoria County as seen in Figure 7. The roadway that was selected was TX-288, a major highway that connects the largest city in Texas to the Gulf coast and its energy sectors. The section selected for data collection spanned 24.6 miles from north of Angleton, TX to south of Freeport, TX, where it merged with TX-36. Each of these sections were collected using two round trips, expanding the total mileage by a factor of four to 98.4 miles. After filtering the dataset for the removal of different surface types, clean images, highly noisy images, and wet/contaminated surfaces, the total image count and mileage would decrease.

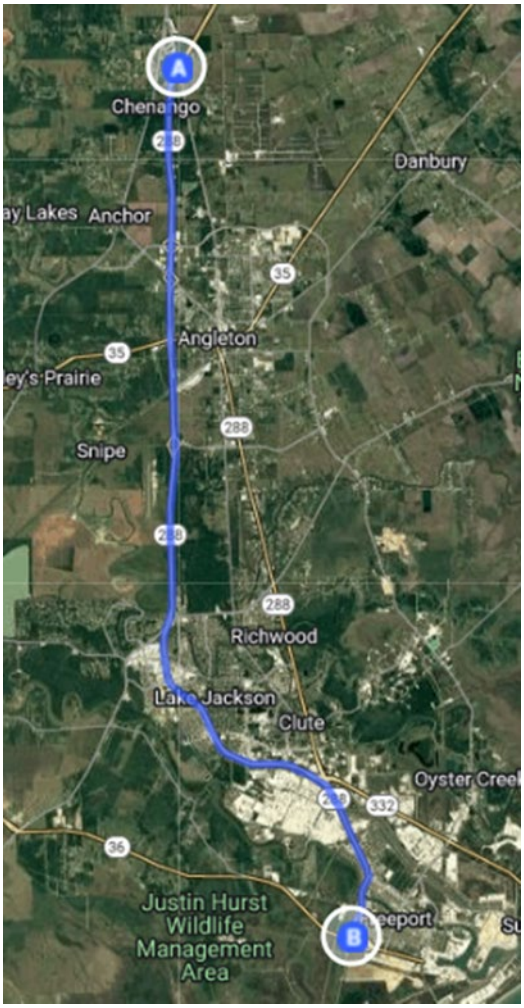


Figure 7. Location of CRCP data collection sites.

The CRCP highway is a massive rural and urban corridor with varying degrees of distress severity. More severe and uncommon distresses are present in this highway with substantial numbers of patches, spalling, transverse joints, and joint failures than in the majority of inland sections.

Asphalt Pavement Dataset

The raw data from the pavement survey vehicle was extracted and converted into two separate image files. These files were pairs of intensity and range images of the road surface, which were then formatted into .bmp files. The images were resized into 4096x2048 rectangular images while retaining the original pixel resolutions of 7mm in the longitudinal direction and 1.042mm in the transverse direction. The resolution frequency at which the images are captured can determine the appearance of distresses, especially with thin hairline cracks. The 3D images were used as the

primary factor for annotating the ground truths, as the 2D images lacked the context and features of depth that were present from the laser sensor. A custom labeling tool was used to create annotation boxes that would locate and label the distresses across the image. The interface of this in-house developed labelling software is shown in Figure 8, where the 3D and 2D images of the road surface are displayed next to each other along with a list of images for a specific section and classifications for the bounding boxes. For asphalt surfaces, a class list consisting of nine labels was used, as displayed in Table 3. Seven of those labels were distresses such as transverse cracking, longitudinal cracking, block cracking, alligator cracking, and failures which could be visualized and determined by the annotators from the 3D images. The other two were non-distress labels, which were added to help the model determine objects on the road surface that were commonly found along the routes the pavement survey vehicle took.

Table 3: Classification list for asphalt pavement surfaces

Distress Items	**Non-Distress Items**
Transverse Cracks	Joints
Longitudinal Cracks	Lane Longitudinal Cracks
Block Cracks	
Alligator Cracks	
Failures	
Sealed Transverse Cracks	
Sealed Longitudinal Cracks	

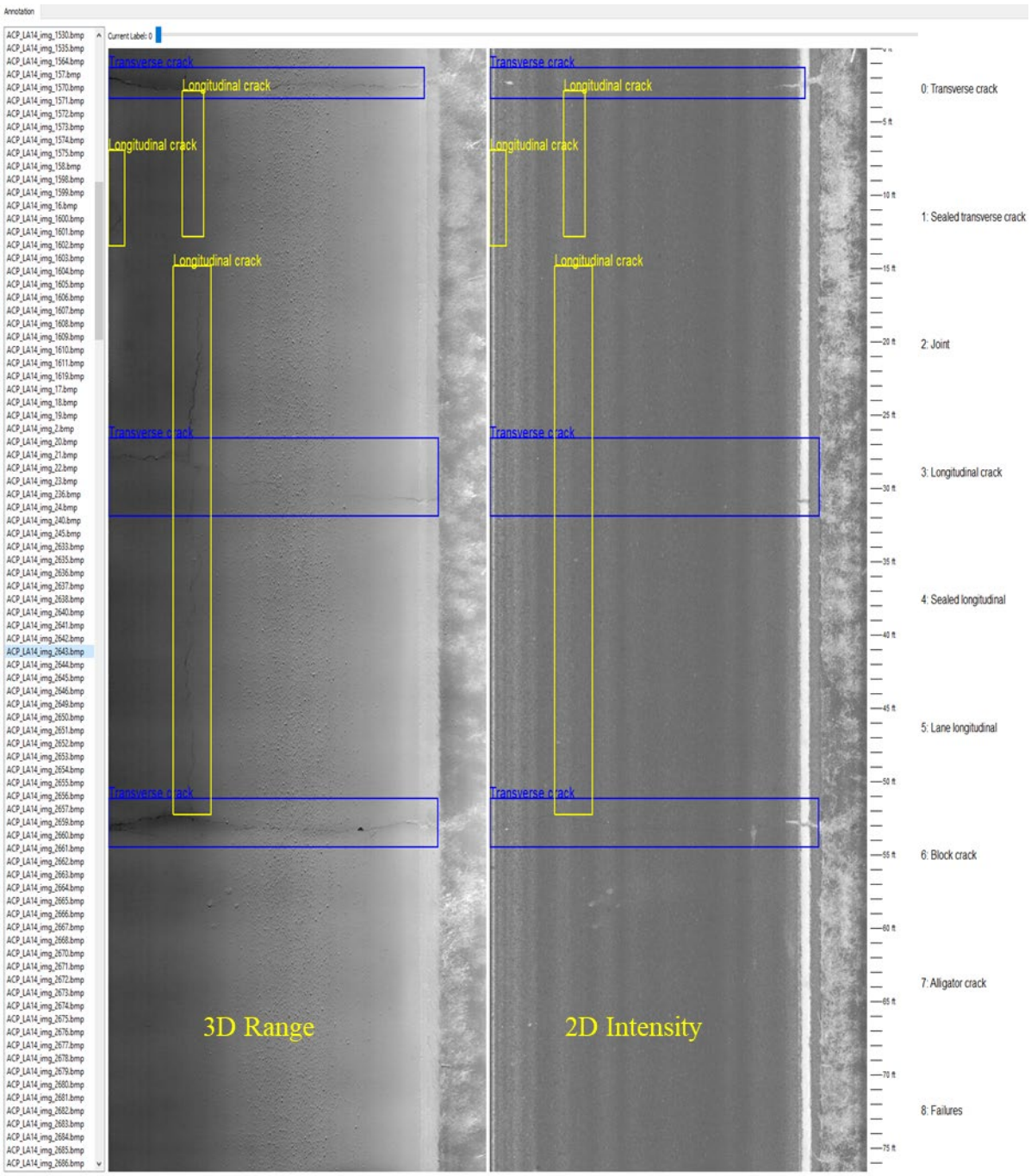


Figure 8. Custom labelling software interface.

To prepare the model for dealing with real-world conditions, the dataset included images with a wide variety of distress severity. Many annotations include thin cracks to avoid any issues with the model failing to detect minor distresses, which could reduce the accuracy of automatic detection. The distresses within the surface area of the pavement within the boundaries of the lane were considered. In the presence of multiple distresses active within a concentrated region of the

image, the types with higher priority and salience were selected for labelling with the annotation boxes. From Figure 9, the event of a distress and non-distress appearing at the same location were labelled accordingly. This meant that if an alligator crack and sealed crack were in the same region, both were annotated with their respective labels. Figure 10 shows the distribution of distresses in both the inland ACP dataset, where 3,855 images have been labeled covering a distance of 34.4 miles, and the coastal ACP dataset, consisting of 2,753 images over 24.5 miles. The most prevalent distresses in both ACP datasets are transverse and longitudinal cracks.

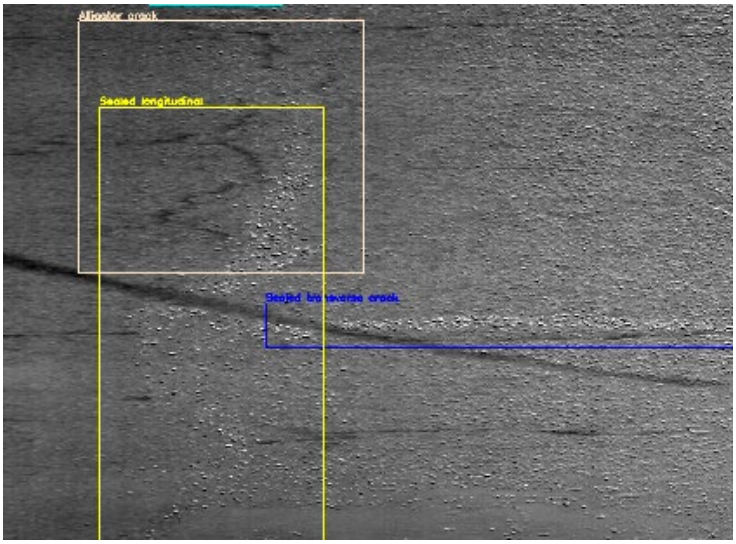


Figure 9: Overlap of distress & non-distress boxes.

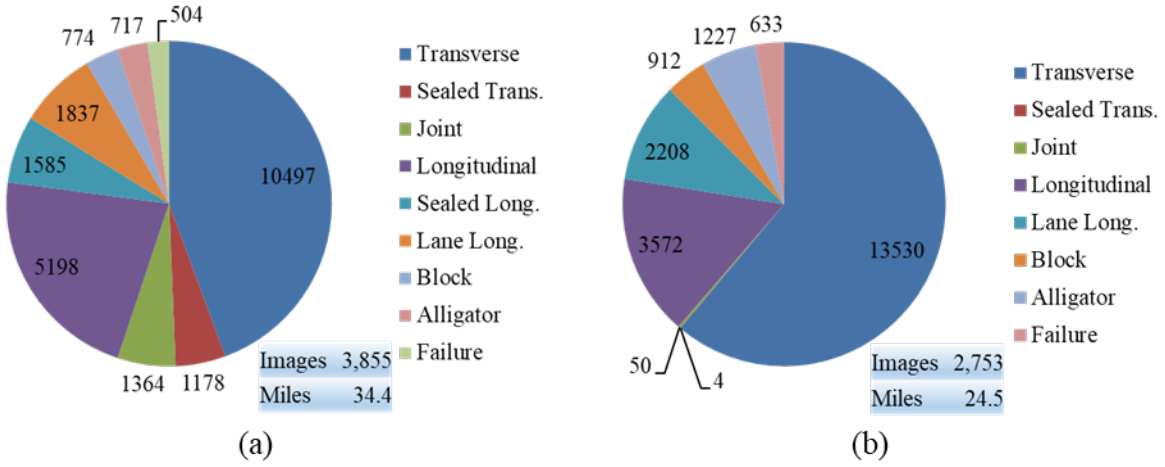
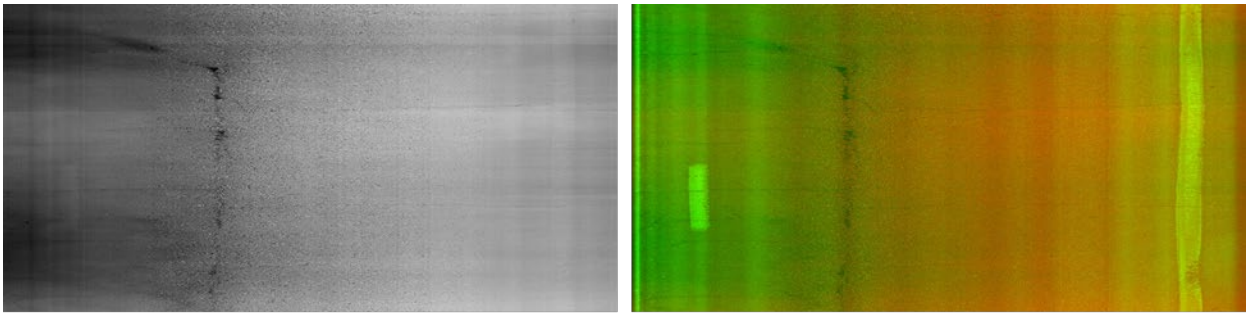


Figure 10. Distress distribution of ACP surface datasets: (a) Inland ACP sections; (b) Coastal ACP sections.

fusion of the 2D intensity and 3D range maps was conducted and evaluated to compare the viability of this approach compared to only using 3D range images as input data. The grayscale image maps were converted to singular color channels with the intensity images laid over the range images to create and RGB image with two occupied and one empty input channel. Figure 11 provides visualization of this technique, where characteristics of both the 2D intensity and 3D range maps are preserved.



(a) (b)

Figure 11. Pavement Surface Images: (a) 3D range image; (b) Fused 2D intensity/3D range image.

Compatibility between the inland and coastal sections was evaluated so that the ability for the model to expand based on location, distress severity, and distress distribution could be determined. Due to the timing of the data collection process, the initial dataset mainly consisted of inland sections for general pavement distress detection.

Rigid Pavement Dataset

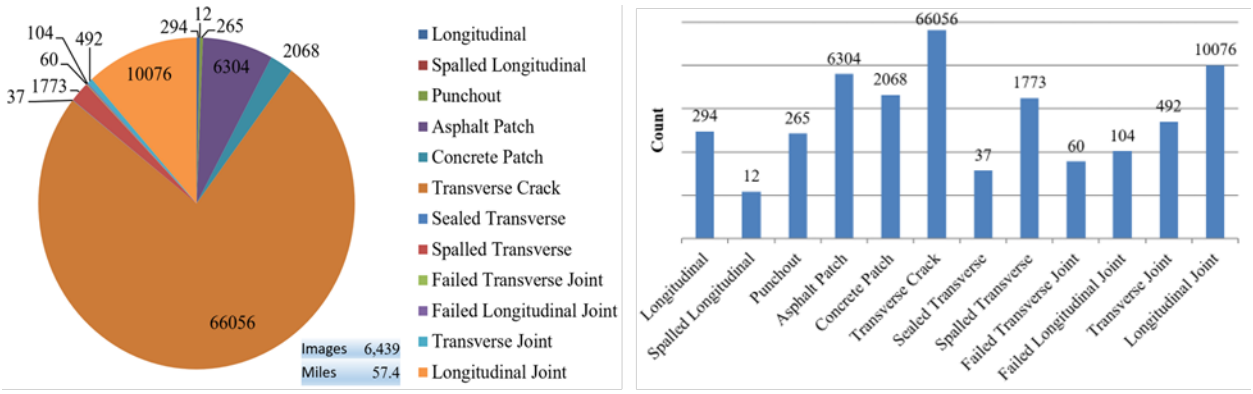
The contrast between the surface characteristics of CRCP and its resilience provides both advantages and disadvantages to the labelling procedure. The resilience of concrete leads to shallower cracks and low occurrences for more severe distresses, causing low visibility in 3D range images. However, the brighter color of the concrete surface allows for higher visibility of cracks, spalling, and patches in the 2D intensity maps. With these two factors in consideration, the labelling process is altered from the initial approach for flexible pavements. Both the 2D and 3D images are used to aid in the labelling and classification of distresses and objects on the rigid surfaces. The most common objects and distresses for CRCP are listed in Table 4 below. As with the class list for asphalt surfaces, both distress and non-distress items are taken into account.

Table 4: Classification list for CRCP dataset

Distress Item	**Common Item**
Transverse Crack	Spalled Longitudinal Crack
Longitudinal Crack	Sealed Transverse Crack
Spalled Transverse Crack	Sealed Longitudinal Crack
Asphalt Patch	Longitudinal Joint
Concrete Patch	Transverse Joint
Punchout	
Failed Transverse Joint	
Failed Longitudinal Joint	

While failed joints are not listed as distresses per the TxDOT pavement rater's manual for CRCP, they are present for jointed concrete pavement (JCP) and do occur often on these sections as well. They can appear either at the edges of concrete patches, bridge approaches, or merging lanes. For longitudinal cracks, spalling occurs with much less frequency due to the design of the pavement type allowing for controlled transverse crack spacing, leading to less longitudinal cracks in general. Although an uncommon form of surface treatment for CRCP, instances for use of crack seals are taken into consideration.

Since the resilience of CRCP along with its high construction and maintenance operations cost do not allow for much opportunity for reliable data capture of surface distresses, TX-288 was the sole section for the rigid pavement dataset. It's location as a key corridor between Houston and the Gulf coast's refineries along with its nearly continuous span of 25 miles provided for optimal conditions to capture multiple distress classes. Figure 12 shows the distress distribution of the CRCP dataset, consisting of 6,439 images that contain 57.4 miles in total. A logarithmic scale was used for the bar graph due to the large disparity in instances between the classes.



(a) (b)

Figure 12. Distress distribution in the CRCP dataset.

The drop in mileage from the data collected stems from the exclusion of images containing non-CRCP surfaces, high levels of noise, heavily imbalanced lighting, low crack visibility in both 2D intensity and 3D range images, and saturated surfaces. The composition of the dataset in Figure 15 may appear to be more favorable to two class types, transverse cracks and longitudinal joints, but those are simply the most common objects available on CRCP surfaces. The design of the reinforced concrete allows for controlled transverse crack spacing to severely limit the appearance and severity of longitudinal cracks. Since this type of rigid pavement allows for continuous reinforcement, the slabs of concrete can extend for miles before a single transverse joint appears while the longitudinal joints serve as boundaries between the lanes. The service TX-288 provides to the region along with the extensive traffic density and loads allowed for the development of more severe distresses such as spalling, punchouts, and patching. While some classes are either not common enough to obtain substantial instances or require the appearance of another class before forming, TX-288 contained ample opportunity to train a model for CRCP sections.

Model Selection

This research opted for the use of CNNs due to their capabilities with visual data and image pixel values. The datasets use fused 2D/3D pavement surface images and require the use of state-of-the-art detection models. From the related work in the literature review, the YOLO series of models were chosen for evaluation. The established use in pavement distress detection of YOLOv5 led to its selection for this study along with YOLOv8 and its subsequent versions up until YOLOv11.

YOLOv5, which was released in 2020, uses cross-stage partial network (CSPNet) in its backbone and Path Aggregation Network (PANet) in its neck to facilitate the balance of speed and accuracy for enhanced real-time detection. Figure 13 displays the architecture of YOLOv5 with further detail.

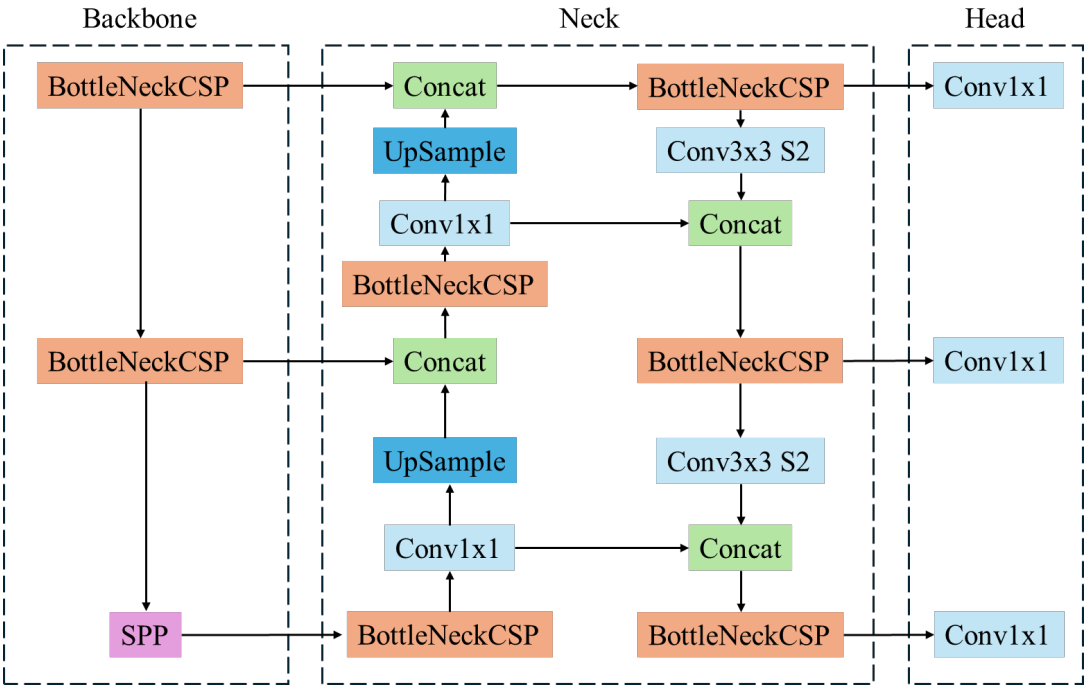


Figure 13. Model architecture of YOLOv5.

The YOLO line of detection models is capable of processing images in a single pass. For YOLOv5, the BottleNeckCSP modules split the image into two parts. Both parts go through convolution with one of them being sent through more convolution, bottleneck, and concatenation layers. This is done in tandem with the spatial pyramid pooling (SPP) node to aggregate the extracted input data into a fixed length for the output data (Jocher *et al.*, 2022). Once that is completed, the information is sent to the prediction heads, where the convolution layers are utilized to determine predicted bounding box coordinates, metric scores, and class label.

With the development of the YOLO models continuing, advancements in the field of deep learning (DL) were made that resulted in enhanced detection capabilities. In YOLOv8, it continued the use of the CSPNet modules in its backbone and improves the model's ability to reduce computational cost by using two-stage feature fusion (C2F) blocks in its architecture, as shown in Figure 14. In addition to this, the SPP block gets upgraded to a Spatial Partial Pyramid-Fast (SPPF)

block in the neck. These changes to the backbone and neck, respectively, allow for YOLOv8 to reduce memory consumption while expanding on the feature fusion functions as well as maintaining the balance of accuracy and computational speed (Jocher, Qiu and Chaurasia, 2023).

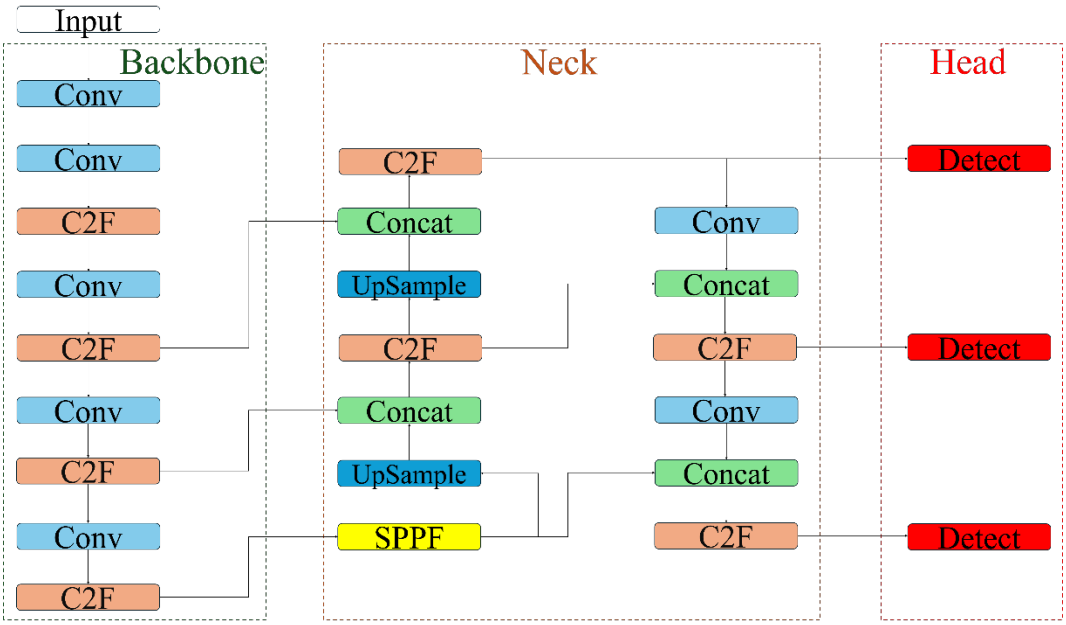


Figure 14. Model architecture of YOLOv8.

Within months of each other, Ultralytics released YOLO versions 9, 10, and 11 in 2024. Despite the short window between releases of each model, these models expand and advance the capabilities of detection for the CNNs. YOLOv9 provides a huge leap in this field with the utilization of Programmable Gradient Information (PGI) and Generalized Efficient Layer Aggregation Network (GELAN). Figure 15 displays the switch from C2F nodes to RepNCSELAN4, an enhanced version of CSP-ELAN architecture, as well as updating the SPPF to SPPELAN to integrate the use of GELAN to optimize accuracy at a slight cost for speed (Wang, Yeh and Liao, 2024). The PGI is utilized in the auxiliary part of the model, resolving a majority of issues with data loss during training.

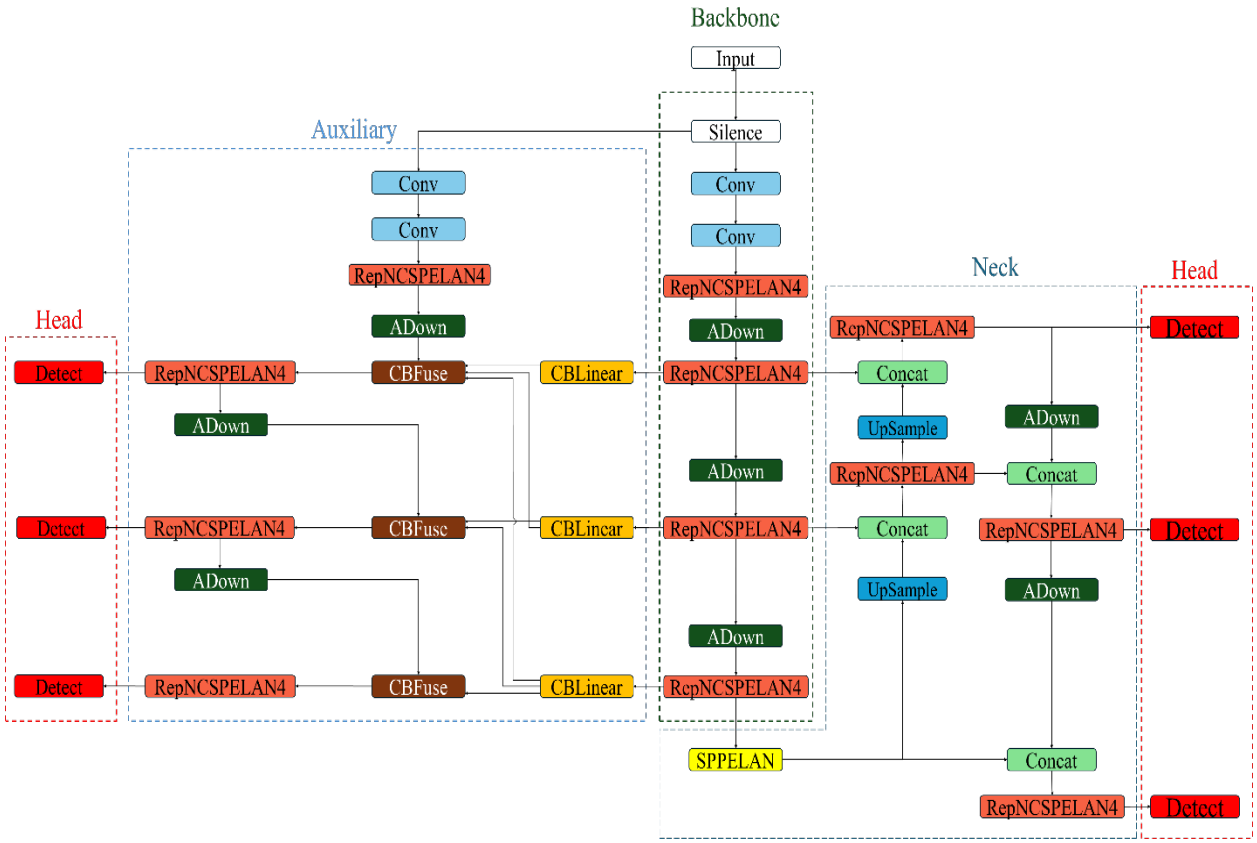


Figure 15. Model architecture of YOLOv9.

In YOLOv10, the backbone utilizes an enhanced CSPNet backbone and PAN neck. Major parts of the advancement from YOLOv9 to YOLOv10 come in the use of non-maximum suppression (NMS)-free training to avoid multiple bounding boxes for the same object. This is done by assigning a single, unique bounding box during the training and inference phases of model development, improving computational efficiency and speed (A. Wang *et al.*, 2024). As for YOLOv11, the continuation of advancing on the C2F blocks is done by using C3K2 blocks and Cross State Partial with Spatial Attention (C2PSA) modules. These blocks use smaller convolution kernels while the C2PSA modules provide focus on important regions within the input image (Jocher and Qiu, 2024). Their architectures can be seen in Figures 16 and 17 below for further detail.

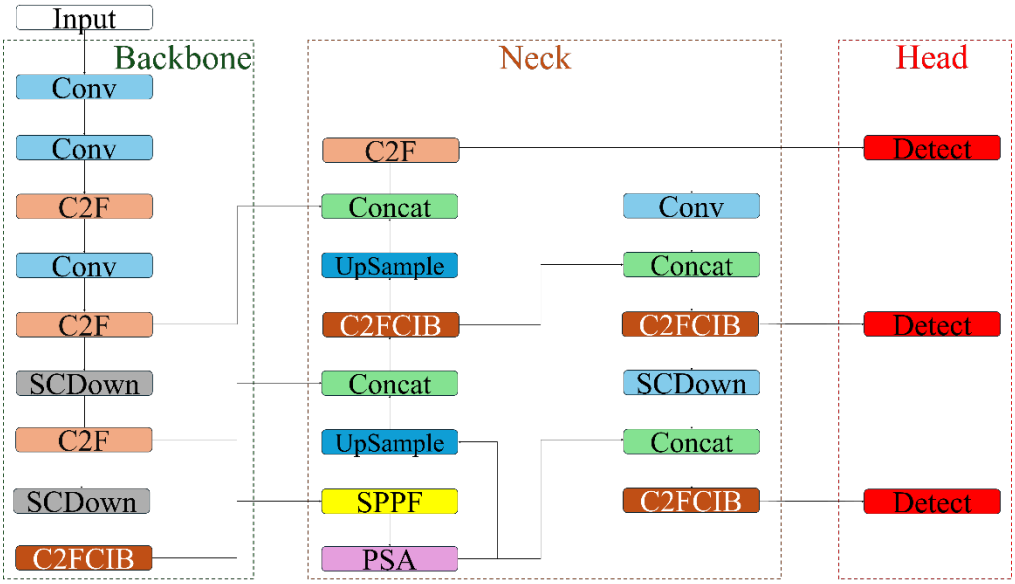


Figure 16. Model architecture of YOLOv10.

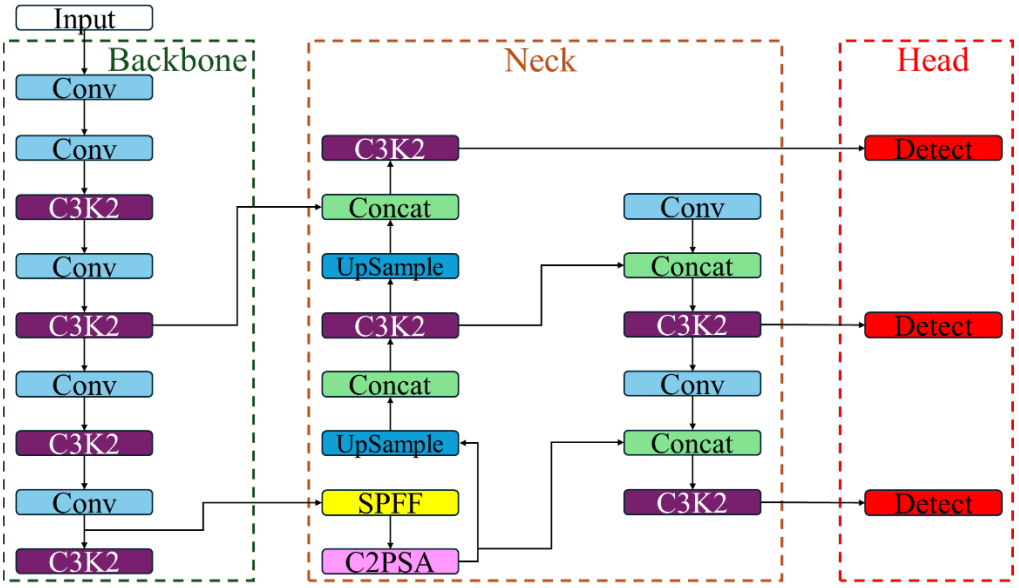


Figure 17. Model architecture of YOLOv11.

Besides using YOLO models, a recent state-of-the-art DL model named Real-Time Detection Transformer (RT-DETR) was considered as a feasible option for the pavement image dataset. It was developed by Baidu to be as efficient as YOLO while attempting to resolve issues faced in feature extraction and inference speed in DL models. As a detection transformer, it bypasses the issues with data loss in NMS while addressing issues in decreased detection speed found in other DETRs. The backbone in Figure 18 is a CNN that extracts the feature maps on an

input image (Zhao *et al.*, 2024). The neck, which is an efficient hybrid encoder, uses an Adaptive Interaction Fusion Integration (AIFI) to fuse the extracted features from the backbone and a Cross-Scale Channel Fusion Fusion (CCFF) to combine the features from several scales.

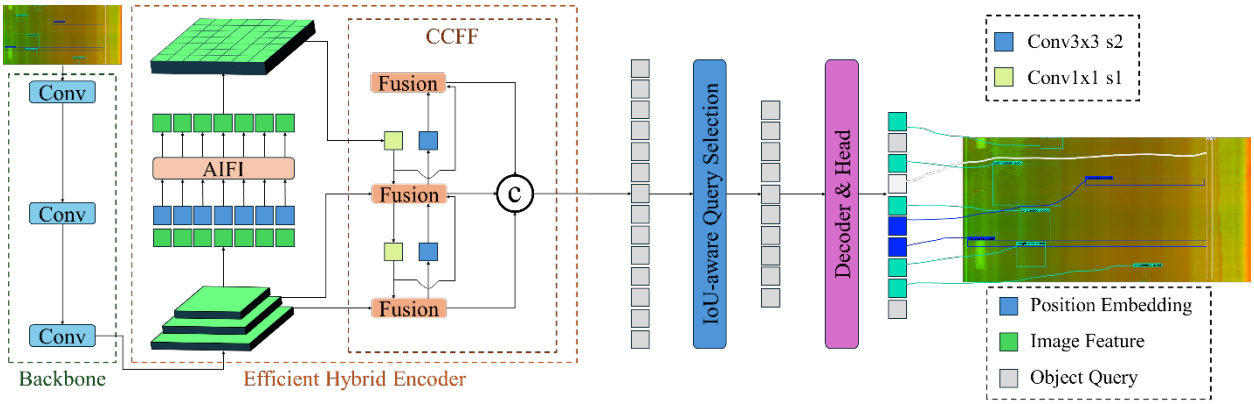


Figure 18. Model architecture of RT-DETR.

These models can all be found in Ultralytics's Github repository, along with their accompanying packages that allow for simple use of model development and training. Since RT-DETR is not a product of Ultralytics, only the large and extra-large sizes of the model are available.

Performance Metrics

As for any AI/ML model, there are key metrics that must be conducted to evaluate the performance of an algorithm. The most common metrics that are used are Equations 5 and 6, precision and recall. Precision is used to quantify a model's success rate for correct predictions. Likewise, recall is used for total predictions from the model across the dataset by taking missed predictions into consideration.

$$Precison = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

In each of these metrics, TP represents true positives, where a predicted detection matches the ground-truth, FP is the false positives, where a prediction is not present in the ground-truth, and FN is the false negative where a prediction was missed by the model when the ground-truth states an object is present (Shalev-Shwartz and Ben-David, 2014). When dealing with more than a single dataset, the intersection-over-union (IoU) in Equation 7 is used when there is overlapping

data between two or more sets. This is also useful for evaluating accuracy of model's predictions with the dataset's ground truth. With a given percentage of overlap, Equation 8 can be used to determine a model's mean average precision (mAP) for its overall performance or individual classes.

$$IoU(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{7}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{8}$$

In Equation 7, A and B are notated as separate datasets. However, detection algorithms utilize predicted labels and ground truths as the inputs for A and B. The amount of overlap between the two labels dictates the threshold for Equation 8, where n is the number of classes in the dataset. The combination of these metrics visualizes the performance and accuracy of most AI/ML models and can alert the presence of errors with the algorithm such as data quality, overfitting, and under-representation of minority classes.

EXPERIMENTAL SETUP

Every model was trained and evaluated with the inland ACP dataset. One model would be selected for further experiments on the coastal ACP and CRCP datasets. The distribution for training, validation, and testing was 8:1:1, respectively. The models were trained for at least 300 epochs, which are complete passes of the images specified for training, with the images being resized from 4096*2048 resolution to 2048*1024. Each training session has a patience value of 100 epochs to end training early once there is no virtual increase in model performance metrics. Each process for the models was evaluated with Equations 5,6, and 9, with a 50% overlap between the labeled and predicted boxes used as a threshold for the mAP score. Three training scenarios were established based on the input image size and data augmentations applied. Scenario 1 uses the unaltered version of the training values, which an input image size of 640x640 and the argumentations found in Table 5. Scenario 2 introduces an increase in image size to 1,024x1,024 as it is the largest square image size available for the resized datasets. Due to the imbalance in the datasets concerning instances of each classification item, the scores are applied to individual classes as well. The orientation of these distresses is the primary characteristic for most classes on

pavement surfaces, leaving little room for additional augmentations to artificially expand the size of the dataset. With scenario 3, all data augmentations in Table 5 are disabled except for the vertical and horizontal flips. Instead, these two data augmentations have their values set to a probability of 0.5 each.

Table 5: Default data augmentations

Augmentation	**Description**	**Default Value**	**Unit**
hsv_h	Alters image hue	0.015	Fraction
hsv_s	Alters image saturation	0.7	Fraction
hsv_v	Alters image brightness	0.4	Fraction
degrees	Rotates image	0.0	Degrees
translate	Translates image	0.1	Fraction
scale	Changes image scale	0.5	Gain
shear	Shifts rows and columns of image	0.0	Degrees
perspective	Changes original image perspective	0.0	Fraction (range of 0–0.001)
flipud	Vertical flip probability of image	0.0	Probability
fliplr	Horizontal flip probability of image	0.5	Probability
mosaic	Resizes and splits four images into one	1.0	Probability

The hue-saturation-value (HSV) color augmentations could reduce the context of the ground truths as both 2D and 3D image types were used for ground truths. For scale, perspective, and shear, these augmentations could affect the detection of distresses with anomalous appearances such as oblique cracks. Translating can cause important regions of the image to be unseen, leading to distresses not being visible enough for proper predictions. For these reasons, the directional flip probability was the most appropriate augmentation that was experimented on.

RESULTS & DISCUSSION

Model Selection Results

In this phase of the study, the six DL models selected for the task of distress detection are evaluated on the inland ACP dataset prior to use on either coastal pavement datasets. Each metric, precision (P), recall (R), and mAP50, for general performance are displayed in Table 6.

Table 6: General model performance on inland ACP dataset.

Models	**Scenario 1**			**Scenario 2**			**Scenario 3**		
	P	R	mAP50	P	R	mAP50	P	R	mAP50
YOLOv5	0.525	0.489	0.49	0.584	0.502	0.545	0.544	0.537	0.526
YOLOv8	0.574	0.507	0.522	0.6	0.544	0.554	0.595	0.528	0.527
YOLOv9	0.565	0.501	0.519	0.64	0.551	0.588	0.583	0.595	0.575
YOLOv10	0.587	0.487	0.495	0.624	0.51	0.533	0.592	0.511	0.532
YOLOv11	0.569	0.495	0.503	0.603	0.546	0.558	0.59	0.545	0.549
RT-DETR	0.63	0.535	0.566	0.603	0.588	0.561	0.662	0.557	0.581

*Note: P – Precision; R - Recall

From these models selected, YOLOv5 scored the lowest while RT-DETR outperformed every YOLO model in terms of detection. This was expected as YOLOv5 is the oldest model used in the study and is not sophisticated as a model such as the later YOLO versions or RT-DETR, the latter of which retains its high detection capabilities as a DETR while achieving relatively higher inference speeds than before. YOLOv8 made visible improvements with its precision and mAP50 scores but lacked in recall compared to YOLOv5. In contrast, YOLOv9 has the most success with its performance metrics. Despite yielding slightly lower precision scores, its recall and mAP50 scores were higher than the other versions of YOLO. YOLOv10's focus to increase processing speed led to the inverse effect of higher precision with lower recall and mAP50 scores. YOLOv11 had similar results to its predecessor but did not exceed it. Inference speed for each model is also provided in Figure 19, with RT-DETR being the slowest model. This is due to its nature as a vision transformer to extract more features for higher accuracy than speed. Consequentially, the complex architecture of YOLOv9 made it the second slowest model. With its GELAN modules and PGI-

integrated auxiliary branch, the slight change in the balance of speed and accuracy found in YOLO models opted for more leverage in the latter.

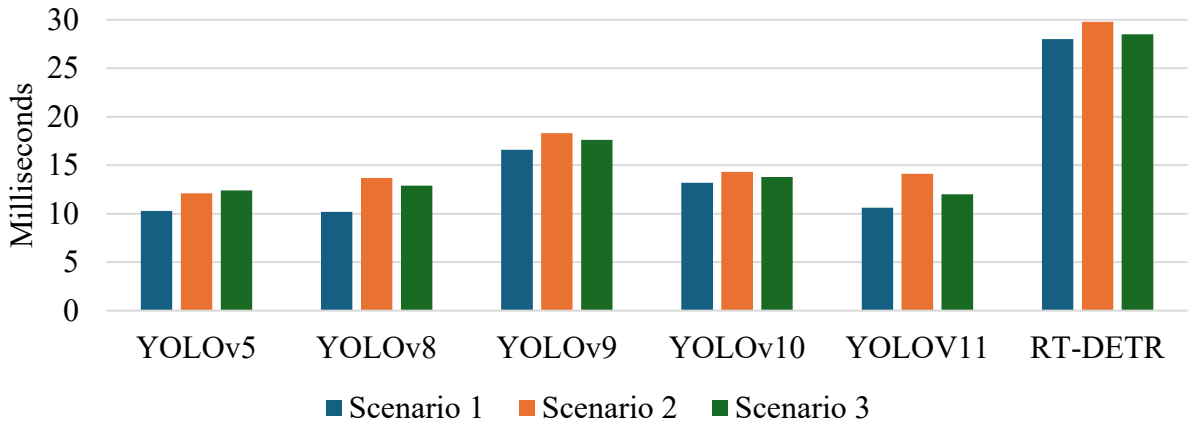


Figure 19. Inference speed of each model.

The average scores of these deep learning models on the inland ACP dataset were not as expected for reliable detection models. While scenarios 2 and 3 did improve the performance metrics to above 0.5 for all models in Table 6, there are still discrepancies it the detection capabilities within the dataset. Figures 20 through 22 displays the metric scores of each model's best-case scenario for all distresses in the classification list. Further investigation into the detection of the individual classes provides context for which distress the models are struggling with.

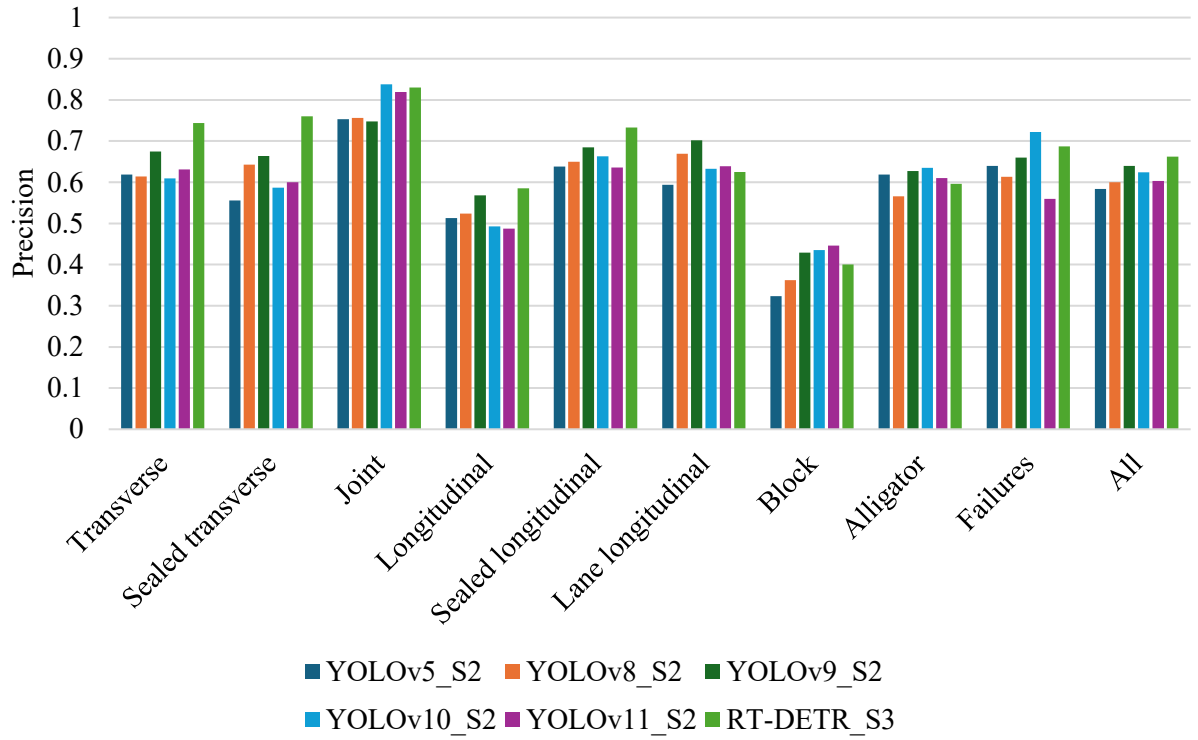


Figure 20. Precision scores for each model on inland ACP dataset.

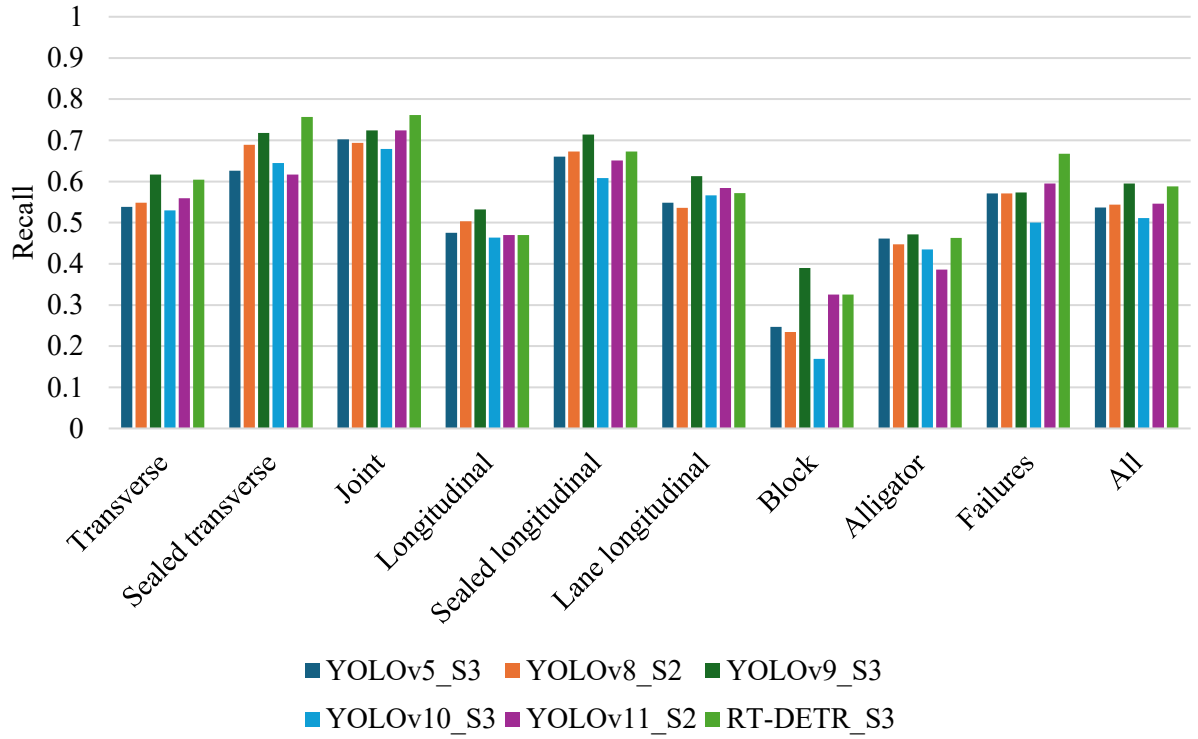


Figure 21. Recall scores for each model on inland ACP dataset.

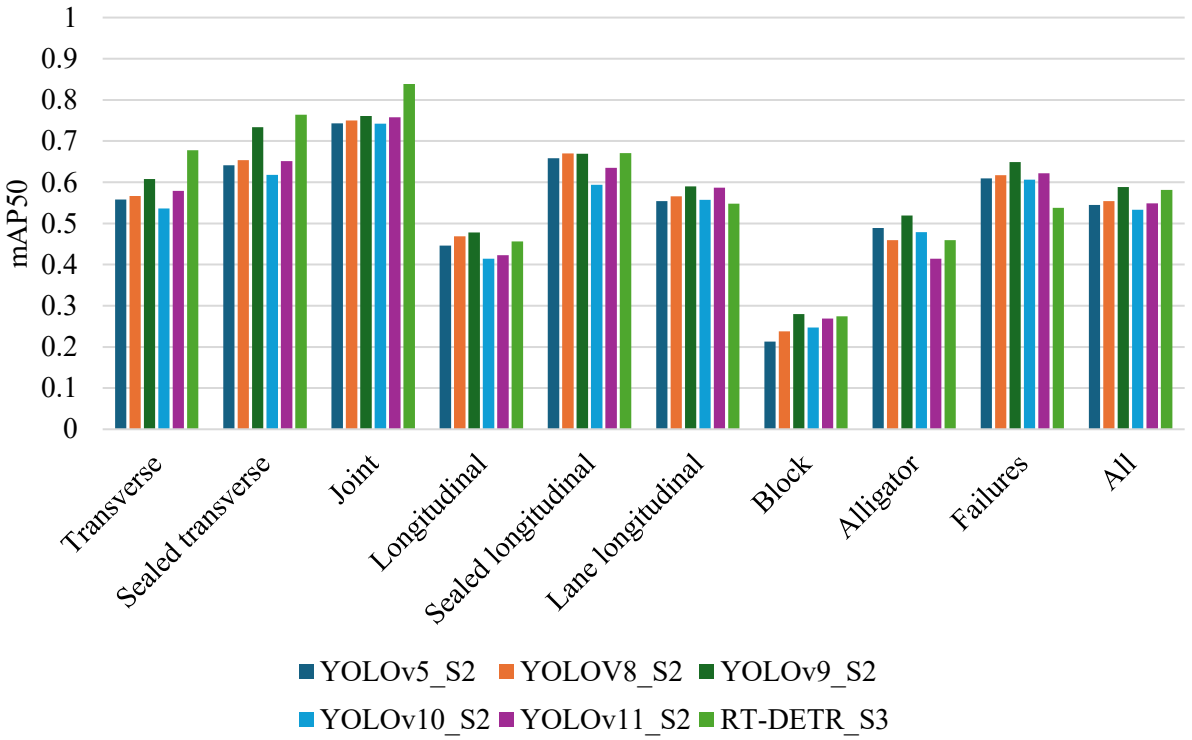


Figure 22. mAP50 scores for each model on inland ACP dataset.

Across nearly every metric for each model, the joints had the highest scores within the selection phase of the study. Their well-defined, linear shape and consistent depth into the ACP surface allowed the DL models to mostly yield scores above 70% with the object class. Sealed cracks and failures were the next highest scoring distresses, scoring above 60% in each metric with the exception of failures in terms of recall. Their appearance in the dataset gave them an advantage over other distresses as sealed cracks would have higher visibility in both image types while failures would present themselves with wider surface areas of darker pixels in 3D images. Transverse and lane longitudinal cracks would obtain precision scores higher than 60% but would only surpass 50% in recall and mAP50. There were several instances of these two distresses in the inland ACP dataset, with transverse cracks having the most labels and lane longitudinal cracks third-most labels of any class.

Longitudinal cracks did not exceed their lane or sealed variants in any of the metrics as their formations were more erratic, much smaller, and not as defined. Between the standard and lane longitudinal crack labels, the latter was more linear, had little deviation in the path of its formation, and were generally found at the edge of the lane. While there were cases where a lane longitudinal

crack would travel across the lane over the course of a handful of images, the cracking was much more uniform and would closely resemble the clean appearance of a joint.

The two remaining area distresses, block and alligator cracking, faced the least success within the inland ACP dataset. Both of these distresses are formed from multiple interconnecting cracks with each of them containing a similar number of instances in the dataset. However, each model would fair much better with detecting alligator cracks. While only reaching above 50% in precision, the DL models would have recall and mAP50 scores above 40% for the alligator cracks. For block cracks, precision scores below 45%, inconsistent recall, and the inability to surpass 30% in mAP50 led it to be the toughest distress for every model in the study. The large discrepancy between these two low-scoring distresses may be due to their strict definitions during labeling. Block cracks must be 1-10 ft^2 while the surface segments of alligator cracking can be below a square foot. In addition to their dimensions, the formations of the two were an influencing factor for successful prediction. Alligator cracks are usually found as webbed cracking in the wheel paths in a section while block cracks require an intersection of two separate transverse and longitudinal cracks. The individual cracks themselves would present issues for detection capabilities as the appearance hairline cracking made it difficult to train a model to detect an object with the width of a few pixels.

Within the individual distresses, a few models would stand out. RT-DETR yielding higher scores across the metrics and class labels made it a strong contender to be chosen as the model for the coastal datasets. Between the YOLO models, YOLOv9 was the strongest choice as its recall scores went unrivaled, maintained a high precision on average, and would only be outperformed on single distress type in terms of mAP50. The metrics between these RT-DETR and YOLOv9 were relatively close, with the former having a slightly bigger margin of success in precision and mAP50. However, its inference speed and computational cost negatively affected its choice as the model to go forward with in this study. YOLOv9, while being the second-slowest DL model in the selection phase, maintained a relatively similar inference speed in Figure 19 to the other models.

Asphalt Pavement Results

From the results during model evaluation, one DL model was chosen. The model that was selected was YOLOv9 due to its detailed architecture using PGI and GELAN to minimize data loss as well as its promising results with improved recall and mAP50 scores. While RT-DETR

performed well and often better in terms of precision and mAP50, its inference speed as recorded in Figure 22 was much slower than that of any of the other YOLO models.

Using the coastal ACP dataset, training was done on YOLOv9 with the same approach as previously mentioned: fused image input data three training scenarios. The difference in class composition and distribution between the coastal and inland datasets stem from treatment types and a singular non-distress item. During the data labelling process, crack seals were hardly encountered beyond a few sealed transverse cracks. This may be due to overlay treatment usage, asphalt patches that cannot be considered as sealed cracks, or lack of maintenance as the distress severity was much higher than observed on the inland sections. The appearance of joints was also not as plentiful with the inland ACP dataset, which may be due to location of where the data was collected. Since the inland data was collected near the I-35 corridor in and around the Austin, TX area, there were more opportunities to encounter municipal roads that had frequent construction boundaries or curbs for pedestrian sidewalks. The coastal dataset, while also travelling through municipalities on occasion, was primarily collected in rural areas on route to the next area of data collection. The performance metrics after training on the coastal data are displayed in Table 7. The increased severity on the asphalt surfaces in the coastal data provided for some notable observations.

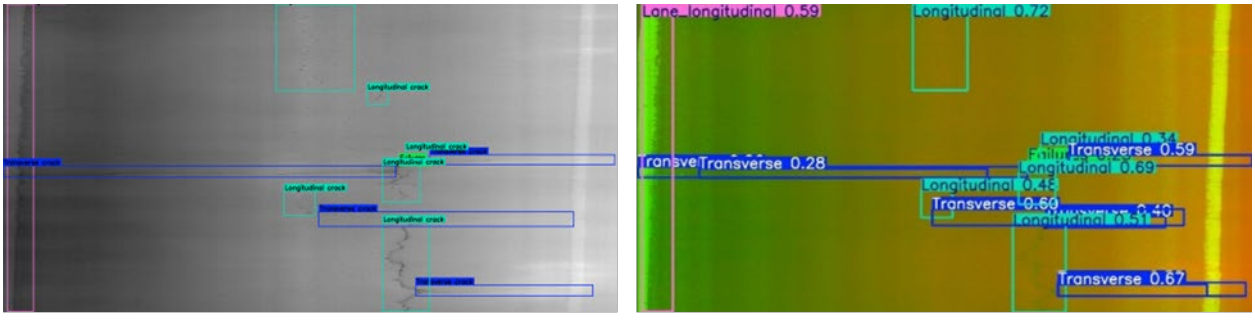
Table 7: YOLOv9 performance metrics on coastal ACP dataset*

Scenarios	**1**			**2**			**3**		
Class	P	R	mAP50	P	R	mAP50	P	R	mAP50
Transverse	0.461	0.623	0.53	0.623	0.555	0.603	0.615	0.618	0.616
Joint	0.996	0.4	0.501	0.653	1	0.928	0.616	1	0.92
Longitudinal	0.431	0.509	0.431	0.553	0.489	0.505	0.55	0.512	0.48
Lane longitudinal	0.485	0.643	0.568	0.62	0.56	0.598	0.644	0.575	0.587
Block	0.399	0.505	0.473	0.521	0.474	0.519	0.533	0.516	0.52
Alligator	0.366	0.433	0.381	0.528	0.413	0.466	0.381	0.375	0.389
Failures	0.31	0.612	0.423	0.509	0.478	0.438	0.57	0.507	0.543
All	0.493	0.532	0.472	0.572	0.567	0.58	0.558	0.586	0.579

*Note: P – Precision; R - Recall

Distress detection on the coastal ACP images fared similar results with inland sections for most of the shared classes. Scenarios 2 and 3 for transverse cracks were mainly over 60% margin due to their common appearances. Joints, despite containing a fraction of the instances available in the inland sections, were fairly detected by YOLOv9 due to their unique characteristics of being entirely linear with clear boundaries in either object width or surface contrast. Longitudinal and lane longitudinal cracks generally had detection rates higher than 50% outside of scenario 1. Alligator cracks and failures did not do as well in terms of detection compared to their inland counterparts, but the most notable result from training on the coastal dataset was major increase in block crack detection. Previously, block cracks would face issues do to their similarity with alligator cracks as they are measured in square footage rather than length along with their composition of 2-3 other class types. These issues caused block cracks to score below 30% mAP50 in each scenario with every model during the selection process. Fortunately, through an extensive peer review process, explicit definitions for each distress, and more instances due to increased distress severity, YOLOv9 would be able to improve its detection rate of block cracks by 26.3%-42.30% across the scenarios.

Despite the fair performance scores in general and for each class, the results were promising in the sense that the dataset did not have as much clarity or quality within the individual images. The prediction capabilities of the standard YOLOv9 model can be seen in Figure 23, where it tended to overpredict compared to the ground truth annotation but was able to correctly identify most of the labeled distresses. The 3D range image with the ground truth labels was used for improved visualization of the ACP distresses.



(a) (b)

Figure 23. Ground truth labels and predicted distresses of YOLOv9 on coastal ACP image: (a) Ground truth labels on 3D range image of ACP surface; (b) Fused image with predicted bounding boxes from YOLOv9.

Brightness affected the visibility of the 3D range maps as the weather during the collection process in Louisiana and Mississippi mainly occurred in clear skies, as compared to cloudy or overcast weather in Texas. The harsh sunlight would often populate the images with noise in the 3D range images, causing blurred edges and surfaces with the surface distresses. The approach to use 2D and 3D images data in tandem helped in bypassing most of the issues with noise, but the surface contrast with asphalt did not aid much unless the discoloration or severe cracking was present. Nonetheless, these difficulties were circumvented as much as possible to return middling results with YOLOv9. If the dataset was much larger with a fairer class distribution, the model would be significantly more adept at detection.

Location Specific Dataset Comparison

As the coastal ACP dataset had lacked the availability of joints and sealed cracks, a two-step process to reinforce YOLOv9's detection abilities on asphalt was conducted. A variant of YOLOv9 that was pretrained on inland ACP data would then be applied to train on coastal data, using the same training scenarios as before. To continue with this method, one singular pretrained model of YOLOv9 was chosen: the YOLOv9 variant trained with scenario 3. The mAP50, precision, and recall scores of the model were adequate enough to outperform most of the models with a convenient processing speed for training. Along with this, scenario 3 exclusively used directional flips as the augmentation technique applied to the dataset, which artificially expanded distress distribution and retained the orientation-based classification of the surface cracks.

Due to the lack of crack seals in the coastal ACP dataset, validation results for those two classes were not available for validation. The results of the YOLOv9 model pretrained on inland ACP data that then underwent training with the coastal ACP data are shown below in Table 8.

Table 8: Performance metrics of pretrained YOLOv9 model on coastal ACP dataset.*

Scenarios	**1**			**2**			**3**		
Class	P	R	mAP50	P	R	mAP50	P	R	mAP50
Transverse	0.672	0.417	0.525	0.619	0.546	0.546	0.6	0.487	0.563
Joint	0.646	0.4	0.482	0.827	0.962	0.962	0.634	0.8	0.798
Longitudinal	0.575	0.435	0.472	0.553	0.537	0.537	0.645	0.5	0.536
Lane longitudinal	0.626	0.52	0.573	0.585	0.556	0.556	0.665	0.567	0.616
Block	0.646	0.461	0.535	0.478	0.568	0.568	0.498	0.347	0.444
Alligator	0.497	0.346	0.388	0.622	0.404	0.404	0.597	0.327	0.434
Failures	0.463	0.388	0.404	0.481	0.582	0.582	0.732	0.507	0.6
All	0.589	0.424	0.483	0.595	0.593	0.593	0.624	0.505	0.57

*Note: P – Precision; R - Recall

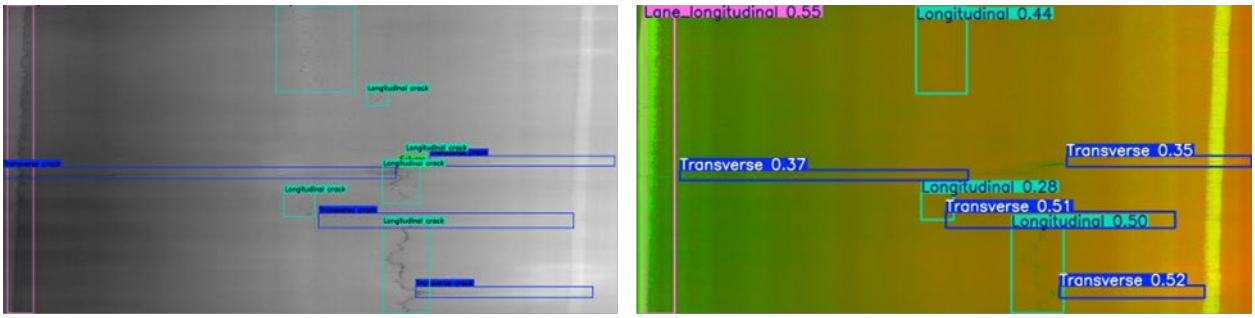
Aside from joints, most of the classes would struggle to surpass 60% accuracy. This is a stark contrast to inland performances, where block cracks were the toughest to detect above 30%. Table 9 provides more context for the effectiveness of this two-step training process, detailing changes in class detection between the results in Tables 6 and 7.

Through the individual scenarios, the two-step process appears to be effective in transfer learning for less available distresses such as joints. While the scores did heavily improve in the first two scenarios, they quickly reverted towards lower scores than usually found in the inland ACP dataset. One of the distress types that did not find much success with the transfer learning applied with the pretrained YOLOv9 model are failures. Compared to the detection of the failure in Figure 23, this version of the model was unable to detect it in Figure 24. This is due to less image samples in the dataset along with irregular distress appearances of the failures, as smaller and shallower failures would contribute to the higher number of instances of the class despite less images.

Table 9: Percent change in performance scores between pretrained and foundation YOLOv9 models on coastal ACP dataset.*

Scenarios	**1**			**2**			**3**		
Class	P	R	mAP50	P	R	mAP50	P	R	mAP50
Transverse	21.1	-20.6	-0.5	-0.4	-0.9	-5.7	-1.5	-13.1	-5.3
Joint	-35	0	-1.9	17.4	-3.8	3.4	1.8	-20	-12.2
Longitudinal	14.4	-7.4	4.1	0	4.8	3.2	9.5	-1.2	5.6
Lane longitudinal	14.1	-12.3	0.5	-3.5	-0.4	-4.2	2.1	-0.8	2.9
Block	24.7	-4.4	6.2	-4.3	9.4	4.9	-3.5	-16.9	-7.6
Alligator	13.1	-8.7	0.7	9.4	-0.9	-6.2	21.6	-4.8	4.5
Failures	15.3	-22.4	-1.9	-2.8	10.4	14.4	16.2	0	5.7
All	9.6	-10.8	1.1	2.3	2.6	1.3	6.6	-8.1	-0.9

*Note: P – Precision; R - Recall



(a) (b)

Figure 24. Ground truth labels and predicted distresses of pretrained YOLOv9 model on coastal ACP image: (a) Ground truth labels on 3D range image of ACP surface; (b) Predicted bounding boxes from pretrained YOLOv9 model on fused 2D/3D image.

This introduction of anomalous distress appearance did have an effect on the detection capabilities of both models. However, at their best performance, the models would not deviate far enough from each other outside of joints and lane longitudinal cracks. With lane longitudinal cracks, aside from clearer definitions for object shape and pattern, they were more prevalent in coastal pavements as they would form as reflective cracks due to the asphalt layer's position above the concrete joints.

Continuously Reinforced Concrete Pavement Results

Despite relying on one sole section for this dataset, the distance, mileage, performance, and image quality of TX-288 would be more than sufficient enough to establish a CRCP dataset. Taking the overwhelmingly high presence of typical objects such as transverse cracks and longitudinal joints into consideration as aspect of the surface type's design, the distribution of both distress and non-distress instances in Figure 12 were bountiful for the most part. As mentioned before, the classes that did not receive sufficient enough labels such as failed joints or spalled longitudinal cracks are either due to their high rarity as a treatment type on CRCP, forming other common distresses, or required additional distresses for its composition or use on the surface. With the caveat of CRCP having greater surface contrast than ACP, the labelling process would rely heavily on a shared use of 2D intensity and 3D range maps as visualization of the pavement surface. For CRCP, it was necessary to fuse the images as the surface did not have issues with contrast as much as ACP has had. The results for the training process are shown in Table 10 below.

Table 10. YOLOv9 performance metrics on CRCP dataset*

Scenarios	**1**			**2**			**3**		
Class	P	R	mAP50	P	R	mAP50	P	R	mAP50
Longitudinal	0.574	0.201	0.243	0.253	0.37	0.245	0.438	0.296	0.27
Spalled Longitudinal	0.642	0.5	0.499	0.443	0.5	0.499	1	0	0.606
Punchout	0.433	0.292	0.293	0.285	0.542	0.343	0.473	0.417	0.39
Asphalt Patch	0.909	0.955	0.974	0.833	0.964	0.977	0.922	0.96	0.981
Concrete Patch	0.782	0.773	0.829	0.541	0.828	0.803	0.776	0.833	0.861
Transverse	0.659	0.737	0.751	0.37	0.743	0.644	0.635	0.614	0.679
Sealed Transverse	0.349	0.25	0.287	0.304	0.25	0.32	0.727	0.5	0.612
Spalled Transverse	0.624	0.667	0.669	0.428	0.826	0.71	0.702	0.713	0.731
Failed Transverse Joint	0.335	0.222	0.346	0.441	0.183	0.181	0.773	0.444	0.519
Failed Longitudinal Joint	0.271	0.182	0.137	0.0886	0.0909	0.152	0	0	0.195
Transverse Joint	0.614	0.585	0.584	0.221	0.391	0.274	0.35	0.34	0.335
Longitudinal Joint	0.893	0.931	0.957	0.723	0.956	0.945	0.904	0.962	0.962
All	0.59	0.525	0.547	0.411	0.554	0.508	0.642	0.505	0.595

*Note: P – Precision; R - Recall

Most of the listed distresses and common objects were able to successfully be detected consistently through the training scenarios. Both asphalt and concrete patches were capable of surpassing 90% along with longitudinal joints while spalled transverse cracks would score above 70% most of the time. Transverse cracks would struggle with staying above 70% but would

consistently float in the 60% range of accuracy. Despite containing over 70% of the total labels in the CRCP dataset, many of the transverse crack were much thinner than their counterparts on the ACP datasets, leading to lower precision scores. Since concrete is much more resilient than asphalt, many cracks would be shallow but still visible on 2D intensity images. Figure 25 contains the ground truths on both the 2D and 3D images in comparison with the predictions by the YOLOv9 model on the fused image since both input images were essential for the labeling process.

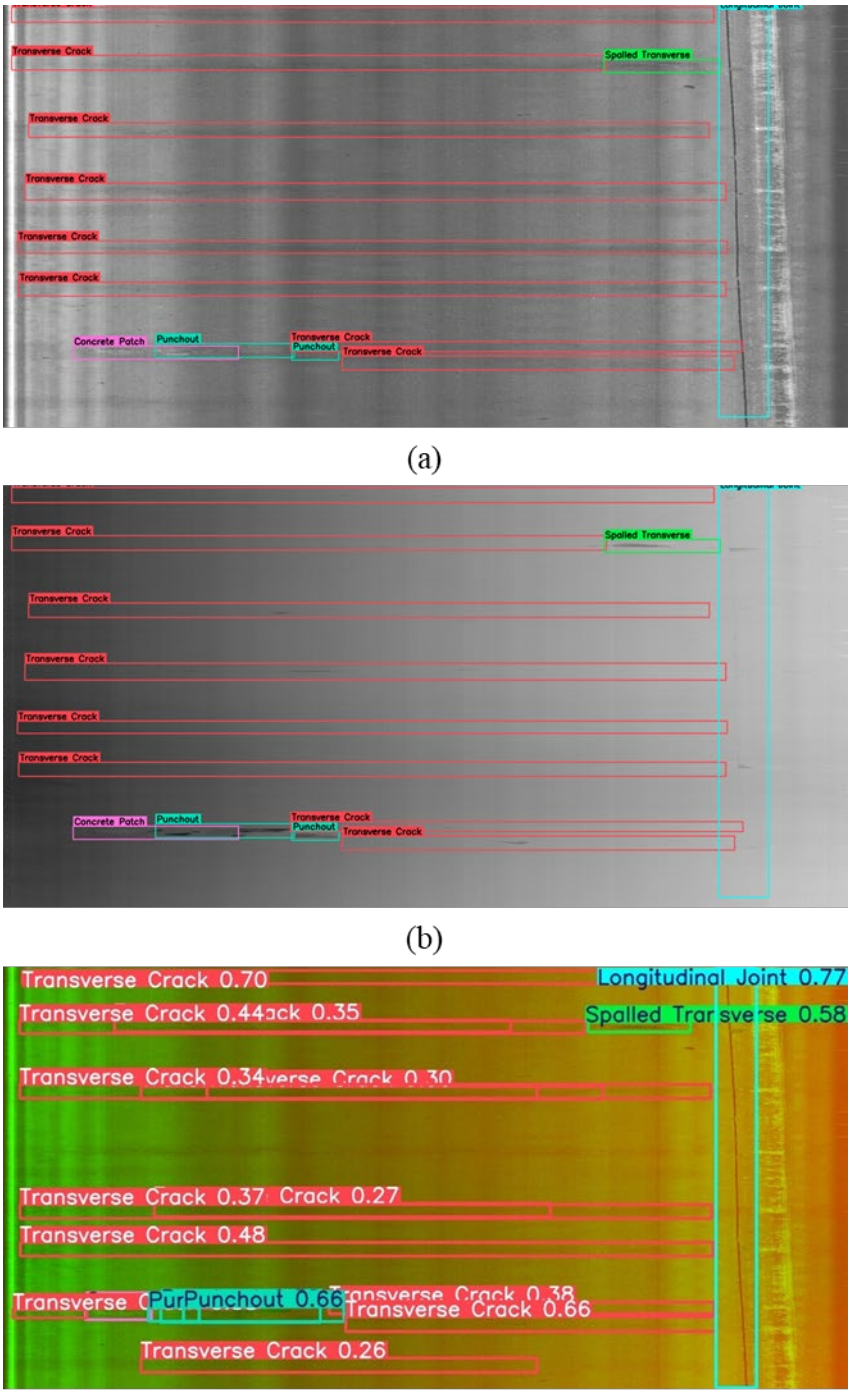


Figure 25. Ground truth labels and predicted distresses of YOLOv9 model on CRCP image: (a) 2D intensity image with ground truth labels; (b) 3D range image with ground truth labels; (c) Fused 2D/3D image with predicted bounding boxes.

Due to the heavy presence of transverse cracks, the punchouts were accurately detected as at least two must be present to fit the requirements for its classification as a distress. However, a

missed detection in the center and false positive prediction at the bottom of the image were caused by extremely thin appearance of the transverse cracks across the dataset.

Punchouts may often get confused with spalled cracks, as the severity of a spall or faulting and presence of two transverse cracks containing the surface voids are necessary for its classification. As transverse cracks on CRCP tend to be extremely thin on several occasions, this guideline becomes harder to register for the model to successfully detect both classes in a consistent manner. This difficult distinction between the two classes led to detection scores lower than 50% for the distress type.

The limited labels and length of the longitudinal cracks had not only reduced the effectiveness of its detection with the model, but it also prevented the appearance of more spalled longitudinal cracks. Existing longitudinal cracks are contained by several transverse cracks when located on the CRCP surface and tend to appear as thin hairline cracks. With low transverse resolution on the image maps, short distance coverage, and lack of common factors associated with consistent causes for longitudinal cracks, YOLOv9 performed poorly with this class. Combined with inconsistent detections on objects that have an insufficient number of instances in total, the average scores ranged from 41.1% to 64.2% for precision, 50.5% to 55.4% for recall, and 54.7% to 59.5% for mAP50.

RECOMMENDATIONS AND CONCLUSIONS

Using three training scenarios on YOLOv5, YOLOv8, YOLOv9, YOLOv10, YOLOv11, and RT-DETR, these models were evaluated for their distress detection capabilities prior to use on coastal pavement images.

As the final candidate, YOLOv9 was chosen to be evaluated on its effectiveness in coastal ACP and CRCP. Despite not having all the listed classification labels, both surfaces were sufficient in containing more dire distresses and higher distress severity. TX-288 as the sole section used for CRCP was more than enough to create a highly diverse dataset for the surface type's distresses and common objects. For the coastal ACP dataset, higher severity would allow for higher instances of severe distresses than the inland asphalt with fewer images.

YOLOv9 had fair results with the coastal ACP as it encountered images with more distress density and noise in its 3D range images from harsh lighting as compared to the weather conditions when the initial inland ACP dataset was collected. The most noteworthy achievement with YOLOv9 on the coastal ACP dataset was its relative success with block cracks, a major change from the findings in the selection phase as the mAP50 scores went from below 30% to above 50% for the model's success rate. When applying transfer learning via a two-step process where the best-performing YOLOv9 model in the inland ACP dataset would then train on the coastal ACP data, the results were not as consistent as expected. While being more efficient with joints, the lack of sealed cracks in general provided little opportunity for the pretrained model to perform well when the goal was to reinforce the accuracy of less represented classes. However, comparing the best performance of the training scenarios conducted on the coastal data for a pretrained variant of YOLOv9 and a foundational YOLOv9 model showed that there were not major differences with the two datasets, providing a chance for compatibility between both regional datasets.

When training on CRCP, YOLOv9 found major success and higher detection rates on the listed distresses for the rigid pavement. Despite imbalance with some of the classes not commonly associated with CRCP and uncommon objects that can appear on these sections, YOLOv9 was able to detect five classes at rates above 60%, four above 70%, and two with mAP50 scores above 90%. As for longitudinal cracks, the pavement's design and distress formation did not allow ample opportunity for the model to effectively detect it.

Using several DL models, ranging from the YOLO series of models and RT-DETR, the detection capabilities of neural networks can be highly effective with the right circumstances. For pavement distress detection and classification, data quality is one of the most important factors as the input data for model training is essential for success. Due to the limited difference and strict guidelines for distress and common road object classification aside from orientation and type of coverage, feature extraction and preservation of distress details is a highly difficult task for training detection models. Recommendations going forward in the field of using ML and DL in pavement distress classification and evaluation can be with either longer and more relevant class types for each pavement surface, two-stage processes with both detection and segmentation models, or stronger, more sophisticated hardware to use the true, full size of the input image without depleting dedicated memory.

The tasks for this research were achieved, as a recent state-of-the-art DL model, YOLOv9, was evaluated and chosen as capable of fairly handling data under real-world conditions such as high levels of noise, distress density, vague distress appearances, and processing images captured at highway speeds. The severity of the coastal ACP dataset provided for a massive turnaround in the toughest class to identify while the approach to transfer learning was viable with varying results. YOLOv9 was able to find higher success rates in CRCP due to the contrast between the distresses and pavement surface. The findings in this thesis provide much needed insight into pavement sections in a region that is exposed to higher rates of declining performances, and how to effectively process the visual data with AI and DL.

DATA AVAILABILITY STATEMENT

All experiment data that support the findings of this study are at the university server site: S:\AA\COSE\ENGR\ENGR-Shares\CETransPMS, and are in the Zenodo Data Share repository site at https://doi.org/10.5281/zenodo.17095715.

REFERENCES

2021 Report Card for America's Infrastructure (2021). American Society of Engineers. Available at: https://www.infrastructurereportcard.org/wp-content/uploads/2020/12/2021-IRC-Executive-Summary.pdf.

Akagic, A. *et al.* (2018) "Pavement crack detection using Otsu thresholding for image segmentation," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija: IEEE, pp. 1092–1097. Available at: https://doi.org/10.23919/mipro.2018.8400199.

Carion, N. *et al.* (2020) "End-to-End Object Detection with Transformers." arXiv. Available at: https://doi.org/10.48550/arXiv.2005.12872.

Cheng, H.D. and Miyojim, M. (1998) "Automatic pavement distress detection system," *Information Sciences*, 108(1–4), pp. 219–240. Available at: https://doi.org/10.1016/s0020-0255(97)10062-7.

Cheng, S. *et al.* (2025) "An Underwater Object Recognition System Based on Improved YOLOv11," *Electronics*, 14(1), p. 201. Available at: https://doi.org/10.3390/electronics14010201.

Choi, P. *et al.* (2020) "Spalling in Continuously Reinforced Concrete Pavement in Texas," *Transportation Research Record: Journal of the Transportation Research Board*, 2674(11), pp. 731–740. Available at: https://doi.org/10.1177/0361198120948509.

Distress Identification Manual for The Long-Term Pavement Performance Program (Fifth Revised Edition) (2014). Federal Highway Administration. Available at: https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/13092/13092.pdf.

Dosovitskiy, A. *et al.* (2021) "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv. Available at: https://doi.org/10.48550/arXiv.2010.11929.

Du, Y. *et al.* (2021) "Pavement distress detection and classification based on YOLO network," *International Journal of Pavement Engineering*, 22(13), pp. 1659–1672. Available at: https://doi.org/10.1080/10298436.2020.1714047.

Elseifi, M.A., Mousa, M.R. and Gaspard, K. (2022) "Impact of the Great Flood of 2016 on the Asphaltic Concrete Road Infrastructure in Louisiana," *Transportation Research Record: Journal of the Transportation Research Board*, 2676(8), pp. 463–474. Available at: https://doi.org/10.1177/03611981221083924.

Eltahan, A.A., Daleiden, J.F. and Simpson, A.L. (1999) "Effectiveness of Maintenance Treatments of Flexible Pavements," *Transportation Research Record: Journal of the Transportation Research Board*, 1680(1), pp. 18–25. Available at: https://doi.org/10.3141/1680-03.

Gong, H. *et al.* (2023) "Automated Pavement Crack Detection with Deep Learning Methods: What Are the Main Factors and How to Improve the Performance?," *Transportation Research Record: Journal of the Transportation Research Board*, 2677(10), pp. 311–323. Available at: https://doi.org/10.1177/03611981231161358.

Guan, J. *et al.* (2021) "Automated pixel-level pavement distress detection based on stereo vision and deep learning," *Automation in Construction*, 129, p. 103788. Available at: https://doi.org/10.1016/j.autcon.2021.103788.

Guo, X. and Wang, N. (2024) "Automated Identification of Pavement Structural Distress Using State-of-the-Art Object Detection Models and Nondestructive Testing," *Journal of Computing in Civil Engineering*, 38(4). Available at: https://doi.org/10.1061/jccee5.cpeng-5864.

He, W. *et al.* (2025) "Object Detection for Medical Image Analysis: Insights from the RT-DETR Model." arXiv. Available at: https://doi.org/10.48550/arXiv.2501.16469.

Hiller, J.E. and Roesler, J.R. (2005) "Determination of Critical Concrete Pavement Fatigue Damage Locations Using Influence Lines," *Journal of Transportation Engineering*, 131(8), pp. 599–607. Available at: https://doi.org/10.1061/(asce)0733-947x(2005)131:8(599).

Hong, F. *et al.* (2023) "Evaluation of Flood-Vulnerable Pavement Network in Support of Resilience in Pavement System Management," *Transportation Research Record: Journal of the Transportation Research Board*, 2677(8), pp. 474–482. Available at: https://doi.org/10.1177/03611981231156934.

Hu, G.X. *et al.* (2021) "Pavement Crack Detection Method Based on Deep Learning Models," *Wireless Communications and Mobile Computing*. Edited by C. Pan, 2021(1). Available at: https://doi.org/10.1155/2021/5573590.

Huang, J., Liu, W. and Sun, X. (2014) "A Pavement Crack Detection Method Combining 2D with 3D Information Based on Dempster-Shafer Theory," *Computer-Aided Civil and Infrastructure Engineering*, 29(4), pp. 299–313. Available at: https://doi.org/10.1111/mice.12041.

Huyan, J. *et al.* (2022) "Pixelwise asphalt concrete pavement crack detection via deep learning-based semantic segmentation method," *Structural Control and Health Monitoring*, 29(8). Available at: https://doi.org/10.1002/stc.2974.

Jocher, G. *et al.* (2022) "ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation." Zenodo. Available at: https://doi.org/10.5281/ZENODO.3908559.

Jocher, G. and Qiu, J. (2024) "Ultralytics YOLOv11." Ultralytics.

Jocher, G., Qiu, J. and Chaurasia, A. (2023) "Ultralytics YOLOv8." Ultralytics.

Karthi, M. *et al.* (2021) "Evolution of YOLO-V5 Algorithm for Object Detection: Automated Detection of Library Books and Performace validation of Dataset," in *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*. *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, Chennai, India: IEEE, pp. 1–6. Available at: https://doi.org/10.1109/icses52305.2021.9633834.

Knott, J.F. *et al.* (2017) "Assessing the Effects of Rising Groundwater from Sea Level Rise on the Service Life of Pavements in Coastal Road Infrastructure," *Transportation Research Record: Journal of the Transportation Research Board*, 2639(1), pp. 1–10. Available at: https://doi.org/10.3141/2639-01.

Li, J. *et al.* (2024) "IDP-YOLOV9: Improvement of Object Detection Model in Severe Weather Scenarios from Drone Perspective," *Applied Sciences*, 14(12), p. 5277. Available at: https://doi.org/10.3390/app14125277.

Li, Y. and Huang, J. (2014) "Safety Impact of Pavement Conditions," *Transportation Research Record: Journal of the Transportation Research Board*, 2455(1), pp. 77–88. Available at: https://doi.org/10.3141/2455-09.

Liang, H. *et al.* (2024) "Automatic Pavement Crack Detection in Multisource Fusion Images Using Similarity and Difference Features," *IEEE Sensors Journal*, 24(5), pp. 5449–5465. Available at: https://doi.org/10.1109/jsen.2023.3267834.

Lin, Z., Wang, H. and Li, S. (2022) "Pavement anomaly detection based on transformer and self-supervised learning," *Automation in Construction*, 143, p. 104544. Available at: https://doi.org/10.1016/j.autcon.2022.104544.

Liu, G. *et al.* (2024) "Pavement Crack Detection Based on Channel Attention Mechanism," in *2024 IEEE Cyber Science and Technology Congress (CyberSciTech)*. *2024 IEEE Cyber Science and Technology Congress (CyberSciTech)*, Boracay Island, Philippines: IEEE, pp. 280–287. Available at: https://doi.org/10.1109/cyberscitech64112.2024.00051.

Liu, J. *et al.* (2020) "Automated pavement crack detection and segmentation based on two-step convolutional neural network," *Computer-Aided Civil and Infrastructure Engineering*, 35(11), pp. 1291–1305. Available at: https://doi.org/10.1111/mice.12622.

Liu, X. *et al.* (2022) "Towards Robust Adaptive Object Detection under Noisy Annotations." arXiv. Available at: https://doi.org/10.48550/arXiv.2204.02620.

Luo, X. *et al.* (2022) "Improving Data Quality of Automated Pavement Condition Data Collection: Summary of State of the Practices of Transportation Agencies and Views of Professionals," *Journal of Transportation Engineering, Part B: Pavements*, 148(3). Available at: https://doi.org/10.1061/jpeodx.0000392.

Ouyang, W. and Xu, B. (2013) "Pavement cracking measurements using 3D laser-scan images," *Measurement Science and Technology*, 24(10), p. 105204. Available at: https://doi.org/10.1088/0957-0233/24/10/105204.

Pavement management information system, rater's manual fiscal year 2021 (2020). Texas Department of Transportation.

Pavement Manual (2021). Texas Department of Transportation. Available at: https://onlinemanuals.txdot.gov/TxDOTOnlineManuals/txdotmanuals/pdm/pdm.pdf.

Prince, S.J.D. (2023) *Understanding Deep Learning*. The MIT Press. Available at: http://udlbook.com.

Qi, F. *et al.* (2024) "Glass Makes Blurs: Learning the Visual Blurriness for Glass Surface Detection," *IEEE Transactions on Industrial Informatics*, 20(4), pp. 6631–6641. Available at: https://doi.org/10.1109/tii.2024.3352232.

Qiu, X. *et al.* (2024) "LD-YOLOv10: A Lightweight Target Detection Algorithm for Drone Scenarios Based on YOLOv10," *Electronics*, 13(16), p. 3269. Available at: https://doi.org/10.3390/electronics13163269.

Quan, Y. *et al.* (2023) "Centralized Feature Pyramid for Object Detection," *IEEE Transactions on Image Processing*, 32, pp. 4341–4354. Available at: https://doi.org/10.1109/tip.2023.3297408.

Rafiei, M.H. and Adeli, H. (2018) "A novel unsupervised deep learning model for global and local health condition assessment of structures," *Engineering Structures*, 156, pp. 598–607. Available at: https://doi.org/10.1016/j.engstruct.2017.10.070.

Ren, J. *et al.* (2022) "Automatic Pavement Crack Detection Fusing Attention Mechanism," *Electronics*, 11(21), p. 3622. Available at: https://doi.org/10.3390/electronics11213622.

Rep. DeFazio, P.A. [D-O.-4 (2021) *H.R.3684 - 117th Congress (2021-2022): Infrastructure Investment and Jobs Act*. Available at: https://www.congress.gov/bill/117th-congress/house-bill/3684 (Accessed: July 10, 2025).

Shalev-Shwartz, S. and Ben-David, S. (2014) *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press. Available at: https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/.

Smola, A. and Vishwanathan, S.V.N. (2010) *Introduction to machine learning*. Cambridge University Press. Available at: https://alex.smola.org/drafts/thebook.pdf.

Sohaib, M., Arif, M. and Kim, J.-M. (2024) "Evaluating YOLO Models for Efficient Crack Detection in Concrete Structures Using Transfer Learning," *Buildings*, 14(12), p. 3928. Available at: https://doi.org/10.3390/buildings14123928.

Sultana, M. *et al.* (2018) "Rutting and Roughness of Flood-Affected Pavements: Literature Review and Deterioration Models," *Journal of Infrastructure Systems*, 24(2). Available at: https://doi.org/10.1061/(asce)is.1943-555x.0000413.

Sutherland, W.J. (1908) “Physiography of the Gulf Coastal Plains,” *Journal of Geography*, 6(11), pp. 337–347. Available at: https://doi.org/10.1080/00221340808985614.

Tang, H. *et al.* (2025) “Divide-and-Conquer: Confluent Triple-Flow Network for RGB-T Salient Object Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(3), pp. 1958–1974. Available at: https://doi.org/10.1109/tpami.2024.3511621.

Tighe, S.L., Ningyuan, L. and Kazmierowski, T. (2008) “Evaluation of Semiautomated and Automated Pavement Distress Collection for Network-Level Pavement Management,” *Transportation Research Record: Journal of the Transportation Research Board*, 2084(1), pp. 11–17. Available at: https://doi.org/10.3141/2084-02.

Tong, Z., Gao, J. and Zhang, H. (2017) “Recognition, location, measurement, and 3D reconstruction of concealed cracks using convolutional neural networks,” *Construction and Building Materials*, 146, pp. 775–787. Available at: https://doi.org/10.1016/j.conbuildmat.2017.04.097.

Varghese, R. and M., S. (2024) “YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness,” in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*. *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, Chennai, India: IEEE, pp. 1–6. Available at: https://doi.org/10.1109/adics58448.2024.10533619.

Voulodimos, A. *et al.* (2018) “Deep Learning for Computer Vision: A Brief Review,” *Computational Intelligence and Neuroscience*, 2018, pp. 1–13. Available at: https://doi.org/10.1155/2018/7068349.

Wang, A. *et al.* (2024) “YOLOv10: Real-Time End-to-End Object Detection.” arXiv. Available at: https://doi.org/10.48550/arXiv.2405.14458.

Wang, C.-Y., Yeh, I.-H. and Liao, H.-Y.M. (2024) “YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information.” arXiv. Available at: https://doi.org/10.48550/arXiv.2402.13616.

Wang, H. *et al.* (2024) “Research on automatic pavement crack identification Based on improved YOLOv8,” *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 18(6), pp. 3773–3783. Available at: https://doi.org/10.1007/s12008-024-01769-3.

Wang, Y., Wang, G. and Mastin, N. (2012) “Costs and Effectiveness of Flexible Pavement Treatments: Experience and Evidence,” *Journal of Performance of Constructed Facilities*, 26(4), pp. 516–525. Available at: https://doi.org/10.1061/(asce)cf.1943-5509.0000253.

Xiong, R. *et al.* (2019) “Performance evaluation of asphalt mixture exposed to dynamic water and chlorine salt erosion,” *Construction and Building Materials*, 201, pp. 121–126. Available at: https://doi.org/10.1016/j.conbuildmat.2018.12.190.

Xu, W. *et al.* (2024) “Violence-YOLO: Enhanced GELAN Algorithm for Violence Detection,” *Applied Sciences*, 14(15), p. 6712. Available at: https://doi.org/10.3390/app14156712.

Yi, H. *et al.* (2024) "Small Object Detection Algorithm Based on Improved YOLOv8 for Remote Sensing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, pp. 1734–1747. Available at: https://doi.org/10.1109/jstars.2023.3339235.

Youwai, S., Chaiyaphat, A. and Chaipetch, P. (2024) "YOLO9tr: a lightweight model for pavement damage detection utilizing a generalized efficient layer aggregation network and attention mechanism," *Journal of Real-Time Image Processing*, 21(5). Available at: https://doi.org/10.1007/s11554-024-01545-2.

Zanevych, Y. *et al.* (2024) "Evaluation of Pothole Detection Performance Using Deep Learning Models Under Low-Light Conditions," *Sustainability*, 16(24), p. 10964. Available at: https://doi.org/10.3390/su162410964.

Zhang, A. *et al.* (2017) "Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces Using a Deep-Learning Network," *Computer-Aided Civil and Infrastructure Engineering*, 32(10), pp. 805–819. Available at: https://doi.org/10.1111/mice.12297.

Zhang, L. *et al.* (2016) "Road crack detection using deep convolutional neural network," in *2016 IEEE International Conference on Image Processing (ICIP)*. *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA: IEEE, pp. 3708–3712. Available at: https://doi.org/10.1109/icip.2016.7533052.

Zhao, M. *et al.* (2022) "Improving the Accuracy of an R-CNN-Based Crack Identification System Using Different Preprocessing Algorithms," *Sensors*, 22(18), p. 7089. Available at: https://doi.org/10.3390/s22187089.

Zhao, Y. *et al.* (2024) "DETRs Beat YOLOs on Real-time Object Detection." arXiv. Available at: https://doi.org/10.48550/arXiv.2304.08069.

Zhou, P. *et al.* (2021) "Study on Performance Damage and Mechanism Analysis of Asphalt under Action of Chloride Salt Erosion," *Materials*, 14(11), p. 3089. Available at: https://doi.org/10.3390/ma14113089.