

# RESEARCH



Report No. UT-25.12

## ASSET CONDITION ASSESSMENT USING AI AND COMPUTER VISION

**Prepared For:**

Utah Department of Transportation  
Research & Innovation Division

**Final Report  
July 2025**

## **DISCLAIMER**

The authors alone are responsible for the preparation and accuracy of the information, data, analysis, discussions, recommendations, and conclusions presented herein. The contents do not necessarily reflect the views, opinions, endorsements, or policies of the Utah Department of Transportation or the U.S. Department of Transportation. The Utah Department of Transportation makes no representation or warranty of any kind, and assumes no liability, therefore.

## **ACKNOWLEDGMENTS**

The authors acknowledge the Utah Department of Transportation (UDOT) for funding this research, and the following individuals from UDOT on the Technical Advisory Committee for helping to guide the research:

- Kevin Nichol
- Benjamin Maughan
- Abdul Wakil
- Sean Berry
- Scott Jones
- Ryan Ferrin
- Clint McCleery
- Chris Whipple
- Kelly Ash

## TECHNICAL REPORT ABSTRACT

1. Report No. UT-25.12	2. Government Accession No. N/A	3. Recipient's Catalog No. N/A	
4. Title and Subtitle Asset Condition Assessment Using AI and Computer Vision		5. Report Date May 2025	
		6. Performing Organization Code	
7. Author(s) Ali Mashhadi, Clancy Black, Erica Wygonik		8. Performing Organization Report No.	
9. Performing Organization Name and Address Wall Consultant Group (WCG), 2139 S 1260 W, Salt Lake City, UT 84119		10. Work Unit No. 5H092 23H	
		11. Contract or Grant No. 25-8370	
12. Sponsoring Agency Name and Address Utah Department of Transportation 4501 South 2700 West P.O. Box 148410 Salt Lake City, UT 84114-8410		13. Type of Report & Period Covered November 2024 to June 2025	
		14. Sponsoring Agency Code PIC No. UT24.204	
15. Supplementary Notes Prepared in cooperation with the Utah Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration			
16. Abstract <p>UDOT faces significant challenges in manually inspecting roadside safety assets across its extensive roadway network, prompting the need for automated condition assessment solutions. This project leverages street-level image data to develop an AI-driven framework for assessing primary assets (guardrails, cable barriers, concrete barriers) and secondary assets (curbs, gutters, retaining walls, shoulder edge drop-offs). Using the YOLO11n model, object detection was evaluated across three settings—multiclass defect detection (mAP@50: 52%, mAP@50:95: 25%), binary classification (mAP@50: 41%, mAP@50:95: 15%), and multiclass with augmentation (mAP@50: 63%, mAP@50:95: 32%)—revealing poor performance despite improved results with augmentation, leading to the adoption of more advanced algorithms. Vision-language models (VLMs) were explored for contextual analysis, with testing conducted to assess both primary and secondary assets, focusing on defect classification, condition rating, and design compliance. Among VLMs, Gemma 3 offers the best accuracy (each image taking about 30 minutes on current GPUs), while Llama 3.2, with its advanced 11B parameter version, delivers the best performance in speed and evaluation. This framework creates significant opportunities to enhance UDOT’s maintenance prioritization, improving roadway safety and infrastructure efficiency.</p>			
17. Key Words Asset Management, Machine Learning, Image Processing, Computer Vision, Barrier		18. Distribution Statement Not restricted. Available through: UDOT Research & Innovation Division 4501 South 2700 West P.O. Box 148410 Salt Lake City, UT 84114-8410 <a href="http://www.udot.utah.gov/go/research">www.udot.utah.gov/go/research</a>	23. Registrant's Seal N/A
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages	22. Price N/A

## TABLE OF CONTENTS

EXECUTIVE SUMMARY .....	7
1.0 INTRODUCTION .....	8
1.1 Overview.....	8
1.2 Objectives .....	9
1.3 Outline .....	10
2.0 LITERATURE REVIEW .....	11
2.1 Asset Detection.....	11
2.1.1 Lane Detection/Road Markings .....	12
2.1.2 Safety Barriers .....	12
2.1.3 Utilities (e.g., Signs, Culverts, Poles) .....	13
2.2 Condition Monitoring .....	14
2.2.1 Road Markings.....	15
2.2.2 Safety Barriers .....	15
2.2.3 Utilities (e.g., Signs, Culverts, Poles) .....	16
2.3 Summary.....	17
3.0 Research Methods.....	19
3.1 Object Detection .....	19
3.2 Multimodal Models.....	20
3.2.1 Zero-Shot Image Classification .....	22
3.2.2 Image-Text-to-Text Models.....	23
4.0 DATA COLLECTION .....	25
4.1 Overview.....	25
4.2 Guardrail Annotation .....	26
5.0 MODEL EVALUATION .....	29
5.1 Object Detection Results .....	29
5.1.1 Multiclass Defect Detection.....	29
5.1.2 Binary Classification (Defect or Not) .....	30
5.1.3 Multiclass with Augmentation .....	31
5.2 Multimodal Deployment.....	34

5.2.1 CLIP .....	35
5.2.2 Llava 1.6 .....	40
5.2.3 QWEN 2.5.....	43
5.2.4 Llama 3.2 .....	51
5.2.5 Gemma 3 .....	65
6.0 CONCLUSIONS.....	70
6.1 Recommendations.....	70
6.2 Limitations .....	71
7.0 REFERENCES .....	73

## LIST OF FIGURES

Figure 1. YOLOv11 Architecture Overview .....	20
Figure 2. Vision Language Models Applications .....	21
Figure 3. VLM Architecture Overview .....	23
Figure 4. Annotated Images within Each Road .....	25
Figure 5. Distribution of Defects in Annotated Images .....	26
Figure 6. Example of an Annotated Image with a Missing/Broken Post.....	28
Figure 7. Example of an Annotated Image with a Flattening .....	28
Figure 8. Example of a Challenging Image for Annotation.....	28
Figure 9. Confusion Matrix for Multiclass Classification .....	30
Figure 10. Confusion Matrix for Binary Classification .....	31
Figure 11. Example of Image Augmentation, Up: Actual Image, Down: Augmented Image .....	32
Figure 12. Example of Image Augmentation, Up: Actual Image, Down: Augmented Image .....	33
Figure 13. Confusion Matrix for Multiclass Classification with Augmentation .....	34
Figure 14. CLIP CASE #1 .....	35
Figure 15. CLIP CASE #2 .....	37
Figure 16. CLIP CASE #3 .....	38
Figure 17. Llava Implementation in Python .....	40
Figure 18. Llava CASE #1 .....	40
Figure 19. Llava CASE #2.....	41
Figure 20. Llava CASE #3.....	42
Figure 21. QWEN Implementation in Python .....	43
Figure 22. QWEN CASE #1, Image of a Spalled Precast Concrete Barrier .....	43
Figure 23. QWEN CASE #2, Image of a Spalled Precast Concrete Barrier .....	44
Figure 24. QWEN CASE #3, Image of a Deflected Guardrail.....	45
Figure 25. QWEN CASE #4, Image of Broken Cable Barrier Posts along a Road.....	47
Figure 26. QWEN CASE #5, Image of Curbs and Gutters in Urban Area.....	48
Figure 27. QWEN CASE #6, Image of Highway Retaining Wall.....	49
Figure 28. Llama Implementation in Python .....	51
Figure 29. Llama CASE #1, Image of a Deflected Guardrail.....	51

Figure 30. Llama CASE #2, Image of a Deflected Guardrail.....	52
Figure 31. Llama CASE #3, Image of a Deflected Guardrail.....	53
Figure 32. Llama CASE #4, Image of Broken Cable Barrier Posts along a Road .....	57
Figure 33. Llama CASE #5, Image of Cast-in-Place Concrete Barriers along a Road .....	58
Figure 34. Llama CASE #6, Mountain View and Concrete Barriers .....	59
Figure 35. Llama CASE #7, Noise Walls .....	60
Figure 36. Llama CASE #8, Fences.....	61
Figure 37. Llama CASE #9, Fences.....	62
Figure 38. Llama CASE #10, Curbs and Gutters.....	63
Figure 39. Llama CASE #11, Precast Concrete Barriers.....	64
Figure 40. GEMMA Implementation in Python.....	65
Figure 41. GEMMA CASE #1, Image of a Deflected Guardrail.....	66
Figure 42. GEMMA CASE #2, Image of a Spalled Precast Concrete Barrier .....	67

**UNIT CONVERSION FACTORS**

**NO UNIT CONVERSIONS**

## **EXECUTIVE SUMMARY**

The Utah Department of Transportation (UDOT) oversees an extensive network of roadways spanning over 6,000 centerline miles. Maintaining the safety and integrity of roadside barriers is critical and challenging due to the scale and complexity of manual inspections. To address this, this project leverages advanced artificial intelligence techniques, utilizing street-level image data to automate defect detection and condition assessment. The evaluation of the YOLO11n model for object detection across three settings reveals poor performance, indicating a need for more advanced approaches. In the multiclass defect detection setting, YOLO11n achieved an mAP@50 of 52% and an mAP@50:95 of 25%, with limited success in differentiating defects such as guardrail deflection and missing or broken posts. The binary classification setting, focusing on defect versus no-defect identification, recorded an mAP@50 of 41% and an mAP@50:95 of 15%, offering an unreliable high-level screening tool with insufficient precision for initial asset triage. The multiclass with augmentation setting, incorporating techniques like mosaic and flip to simulate complex scenes and increase data diversity, improved performance to an mAP@50 of 63% and an mAP@50:95 of 32%, yet these results remain unpromising, underscoring the limitations of YOLO11n despite augmentation's role in addressing the limited defect dataset and varied environmental conditions. It is also believed that the angle of the images and the distance of the barrier from the camera position limits the model's ability to achieve accurate results. Consequently, the research team shifted focus to more advanced and groundbreaking algorithms.

This project tested vision-language models (VLMs) to assess both primary assets (guardrails, cable barriers, concrete barriers) and secondary assets (curb and gutter, retaining walls, fences), leveraging their contextual analysis capabilities to complement object detection. VLMs are explored for tasks such as defect classification (e.g., distinguishing rust from spalling), condition rating through text prompts like "severely damaged curb," and design compliance checks against standards like the Manual for Assessing Safety Hardware (MASH). Among the VLMs, Gemma 3 offers the best accuracy (though each image takes about 30 minutes on current GPUs), while Llama 3.2, with its advanced 11B parameter version, delivers the best performance in terms of speed and evaluation. This evaluation integrates VLM insights with prior outputs, enhancing the accuracy and interpretability of the asset management framework and supporting real-time monitoring and broader infrastructure assessment across Utah's roadway network.

## **1.0 INTRODUCTION**

### **1.1 Overview**

The U.S. transportation network, valued at over \$25 trillion, relies on critical infrastructure assets like pavements, bridges, and safety barriers to support economic activity and public safety. Nationally, one-third of motor vehicle crash fatalities (~13,600 of 40,990 in 2023) result from run-off-road incidents, with 40–60% linked to driver inattention or fatigue, underscoring the importance of barriers. However, 36% of U.S. bridges require major repair or replacement, and aging infrastructure contributes to 44% of unscheduled downtimes, costing \$50 billion annually. Maintenance consumes 43% of the transportation budget, with state and local governments spending \$154 billion on highways and roads in 2021, yet funding lags behind needs due to declining fuel tax revenues. Regular maintenance, recommended every 1–2 years for barriers to address corrosion or structural damage, is essential to ensure safety and functionality. In Utah, safety barriers—guardrails, cable barriers, cast-in-place, and precast concrete barriers, are vital for preventing collisions and enhancing roadside safety. Valued at \$535 million and classified as Tier 2 assets in UDOT’s 2023 Transportation Asset Management Plan (TAMP), these barriers span thousands of miles across the state’s roadways. Environmental factors, crashes, and vandalism degrade these assets, necessitating robust maintenance strategies.

Manual inspection of these assets is a labor-intensive and costly process, often leading to delays in identifying and addressing deficiencies. The sheer volume of barriers across Utah’s roadways makes comprehensive manual assessments impractical, increasing the risk of undetected issues that could contribute to crash severity. To address these challenges, this research project leverages advanced computer vision and artificial intelligence (AI) technologies, specifically convolutional neural networks (CNNs), to develop an automated asset condition assessment model. By utilizing images collected by UDOT’s contractor, Pathway Services, this project aims to streamline the evaluation process, providing accurate, georeferenced condition data for barriers and exploring the feasibility of extending this approach to secondary assets like retaining walls, curb and gutter, and fences. This report presents the progress made in the literature review, data collection, image annotation, model development, and validation, laying the foundation for a transformative

approach to infrastructure management that enhances efficiency, reduces costs, and improves the safety and reliability of Utah’s transportation network.

## 1.2 Objectives

The primary goal of this research project is to enhance the efficiency, accuracy, and safety of UDOT’s transportation asset management by developing advanced methods for condition assessment of roadside safety assets. The specific objectives are as follows:

1. **Develop a Computer Vision Model for Object Detection:** Utilize convolutional neural network (CNN)-based object detection models, such as YOLO and Faster R-CNN, to identify problematic locations across primary assets—guardrails, cable barriers, and concrete barriers (cast-in-place and precast)—using Pathway Services image data. The model will detect defects such as guardrail deflection or leaning.
2. **Explore Vision-Language Model (VLM) Applicability for Primary Asset Condition Evaluation:** Investigate the feasibility of applying VLMs to assess the condition of primary assets by analyzing image-text interactions. This objective aims to evaluate whether VLMs can accurately classify defects (e.g., rust, deformation, or cracking) in guardrails, cable barriers, and concrete barriers, enhancing the robustness of condition assessments beyond traditional object detection methods.
3. **Explore VLM Applicability for Secondary Asset Condition Evaluation:** Assess the potential of VLMs to evaluate the condition of secondary assets, including curb and gutter, retaining walls, and fences. This objective seeks to determine if VLMs can detect relevant defects, such as cracks in curbs or leaning fences, to support an asset condition assessment.

These objectives aim to deliver a scalable, automated solution for UDOT’s barrier condition assessments, improving maintenance prioritization, reducing costs, and enhancing roadway safety, with potential applications for a wider range of transportation assets.

### 1.3 Outline

This report is structured to comprehensively document the development, implementation, and outcomes of the AI-driven barrier condition assessment project for UDOT. The organization is as follows:

- **Introduction:** Establishes the context for the project, highlighting the significance of safety barriers (guardrails, cable barriers, and concrete barriers) in Utah and nationally, the limitations of manual inspection methods, the objectives to develop computer vision and vision-language models (VLMs) for condition assessment, and the report's structure.
- **Literature Review:** Summarizes existing research on AI and computer vision applications in transportation asset management, focusing on object detection models (e.g., YOLO, Faster R-CNN) for barrier assessment and the potential of VLMs for evaluating primary and secondary assets, providing a foundation for the project's methodology.
- **Research Methods:** Details the methodologies, including development of convolutional neural network (CNN)-based object detection models to identify defects in primary assets (guardrails, cable barriers, concrete barriers) and the exploration of VLMs for condition assessment of both primary and secondary assets (curb and gutter, walls, fences).
- **Data Collection:** Describes the collection of Pathway Services' street-level image data, covering Utah's primary and secondary assets, the annotation process for training AI models, and challenges encountered during data collection.
- **Model Evaluation:** Analyzes the performance of the AI models in detecting defects, such as guardrail deflection. This section assesses model accuracy, compares automated results with manual assessments, and evaluates VLM applicability for both asset types.
- **Conclusion:** Summarizes the project's findings, the effectiveness of the AI-driven approach in improving UDOT's asset management, recommendations for integrating automated assessments into maintenance practices, and potential future applications for secondary assets.

## **2.0 LITERATURE REVIEW**

### **2.1 Asset Detection**

In recent years, computer vision techniques, particularly deep learning frameworks, have been extensively applied to detect and classify various road infrastructure assets, including safety barriers, road lanes, markings, and utilities such as culverts and poles. These assets are vital to ensuring road safety and facilitating routine maintenance, making their accurate detection a fundamental task in transportation infrastructure management. Several studies have demonstrated the applicability of deep learning and computer vision in the detection of these assets.

A significant body of research has focused on detecting safety guardrails in construction sites, where accurate and timely detection is essential for mitigating safety risks. For instance, Peinado et al. (2023) explored the use of Unmanned Aerial Systems (UAS) combined with deep learning algorithms to inspect temporary guardrails on construction sites. The system aims to detect missing or failed guardrails using aerial imagery processed with machine learning models, offering a cost-effective approach to improve safety management during critical construction phases. Similarly, Kolar et al. (2018) applied deep learning, specifically transfer learning with the VGG-16 model, to detect safety guardrails in 2D images. Their CNN-based model achieved a high accuracy of 96.5%, highlighting the potential of deep learning in enhancing the detection and safety of guardrails on construction sites, particularly in preventing falls from heights.

Beyond construction sites, asset detection also extends to transportation management systems. (Strain et al., 2020) developed a decision support system that integrates computer vision with GPS-enabled inertial measurement units (IMU) to automatically verify and update highway asset inventories. Tested on the A27 highway in the UK, the system achieved a 91% accuracy in detecting assets such as traffic signs and marker posts, underscoring the utility of combining camera imagery with location data for effective asset monitoring. The following subsection will further review the literature on using computer vision in different areas of transportation.

### 2.1.1 Lane Detection/Road Markings

Lane and road marking detection has received extensive research attention, especially in the context of autonomous driving and Advanced Driver Assistance Systems (ADAS). For lane detection, traditional methods like Hough transforms and Canny edge detection are often combined with deep learning techniques like convolutional neural networks (CNNs) for higher accuracy under challenging scenarios like worn-out markings or occlusion (Mamun et al., 2022). A common approach is to use CNNs for detecting lane boundaries, which can be particularly useful for tracking lane changes and ensuring driver safety. Additionally, recent studies emphasize the use of LiDAR data alongside vision-based systems to enhance detection accuracy, especially in urban environments (Xie et al., 2023).

### 2.1.2 Safety Barriers

Safety barriers, including guardrails and crash barriers, play a critical role in preventing serious injuries when a vehicle leaves the roadway, especially in high-risk areas. Computer vision systems have become increasingly sophisticated in detecting and tracking these barriers using various object detection techniques, such as Convolutional Neural Networks (CNNs) and edge-detection algorithms. These systems can analyze images and point cloud data to identify safety barriers even in complex environments with varying lighting conditions, shadows, or occlusions (S. Xu et al., 2021). Deep learning models like YOLO and Faster R-CNN are particularly effective in identifying these barriers across different environmental contexts, improving both detection speed and accuracy (Huang et al., 2009).

Several studies have advanced the use of LiDAR and other vehicle sensor data for detecting guardrails in real time, offering promising applications for autonomous driving. (Gumaelius, 2022) proposed two approaches—one using semantically labeled LiDAR data and another using raw LiDAR data—to detect and vectorize guardrails for landmark-based localization. Both methods demonstrated strong performance, with the raw LiDAR approach proving more robust in detecting longer sections of guardrails, offering potential solutions for real-time autonomous vehicle localization in areas with poor satellite signal coverage. Deep learning methods have also been applied to more complex environments where vegetation, vehicles, and other obstacles obscure

safety barriers. For example, (H. Jin et al., 2021) developed a Mask R-CNN algorithm for guardrail detection, achieving a high average precision of 94.38% and a recall rate of 93.8% in diverse environments, making the model reliable for road management and maintenance applications. Other studies have explored the integration of image and point cloud data for improved guardrail detection accuracy. (Matsumoto et al., 2021) demonstrated that combining image features with point cloud data collected via mobile mapping systems (MMS) significantly enhances the detection and classification of guardrails, further enabling their application in asset management and 3D mapping. Similarly, Vidal et al. (2020) used mobile laser scanning data to segment and classify barriers, achieving better performance for guardrail detection compared to concrete barriers.

Traditional computer vision approaches also offer valuable insights. For instance, (Z. Xu et al., 2020) applied various image segmentation algorithms, such as active contour and Graph Cut, to detect guardrails from mobile video data captured by patrol cars. Their improved Lazy Snapping algorithm proved effective in segmenting guardrails across diverse environments, offering a practical solution for road maintenance. Additionally, Scharwächter et al. (2014) introduced a novel stereo camera-based method for detecting guardrails in highway settings, leveraging geometric and appearance cues to classify barriers with high accuracy, particularly benefiting Advanced Driver Assistance Systems (ADAS).

### 2.1.3 Utilities (e.g., Signs, Culverts, Poles)

Utilities such as culverts, poles, and drainage systems are essential for ensuring road performance and safety, particularly in managing stormwater and preventing damage. Computer vision systems, leveraging techniques like image segmentation and object detection, have been applied to detect these assets efficiently. The detection of culverts and other subsurface utilities often involves the use of aerial or drone imagery combined with advanced computer vision algorithms, allowing for accurate mapping and condition assessment (Huang et al., 2009). In complex environments where traditional methods struggle due to occlusions, integrating LiDAR data with deep learning models significantly improves detection accuracy, enabling better performance in recognizing these critical utilities (S. Xu et al., 2021).

The integration of computer vision techniques with Geographic Information Systems (GIS) has further enhanced large-scale monitoring of road infrastructure. By streamlining the collection and analysis of data across diverse environments, this integration reduces manual efforts and enables real-time detection and condition assessment of road assets. This advancement improves the management of transportation infrastructure, offering a scalable solution for asset inventory and maintenance (Xie et al., 2023). In addition to utilities, advancements in deep learning have improved the detection of other road assets. For instance, Yu et al. (2022) developed SignHRNet, a deep learning architecture that accurately detects street-level traffic signs even in challenging environments. Using high-resolution networks and attention mechanisms, the model efficiently processes occluded and deformed signs, showing promise for intelligent transportation and autonomous driving systems.

In construction site safety, computer vision has been employed to automate the detection of missing barricades. (Chian et al., 2021) compared two approaches—Masks Comparison Approach (MCA) and Missing Object Detection Approach (MODA)—with MODA demonstrating superior performance in identifying missing safety barriers, which is critical for reducing fall-related accidents on construction sites. Moreover, video-based detection systems have been developed for road asset management. For example, Balali & Golparvar-Fard (2015) proposed a scalable image parsing method that segments and recognizes various roadway assets, such as traffic signs, traffic lights, and pavement markings, from car-mounted video streams. Their method showed strong accuracy, making it a practical tool for large-scale asset management. Additionally, their later work (Balali & Golparvar-Fard, 2016) evaluated computer vision algorithms for traffic sign detection, highlighting the effectiveness of combining histogram of oriented gradients (HOG) with support vector machines (SVM) for accurate asset classification.

## **2.2 Condition Monitoring**

Condition monitoring plays a crucial role in ensuring the safety, functionality, and compliance of road infrastructure assets such as safety barriers, road lanes, road markings, utilities, and culverts. Computer vision-based systems, combined with LiDAR and other advanced technologies, are particularly effective for automating real-time assessments of these assets. The

assessment process can be consolidated under one category to address both Health Evaluation (e.g., evaluating the current physical state of an asset) and Adequacy/Design Evaluation (e.g., determining whether an asset is functioning according to design specifications and meeting the necessary regulatory or safety standards), ensuring a holistic review of the physical condition and design standards compliance of these infrastructure components.

### 2.2.1 Road Markings

Road markings are critical for guiding traffic and ensuring road safety. Over time, these markings deteriorate due to weather conditions and heavy traffic, affecting their visibility and clarity. Computer vision systems have been deployed to evaluate the health of pavement markings by detecting faded or missing sections. For instance, Wang et al. (2021) developed a model that calculates the degree of wear in pavement markings and prioritizes repainting efforts based on condition assessments. In addition, adequacy evaluations check if road markings meet visibility and placement standards, ensuring compliance with traffic flow regulations and safety guidelines (Mamun et al., 2022).

### 2.2.2 Safety Barriers

Safety barriers, such as guardrails, are highly susceptible to wear, corrosion, and damage caused by vehicle collisions. Vision-based condition monitoring systems leverage object detection algorithms and deep learning models to automatically identify and assess the structural integrity of these barriers. These models detect physical damage, such as cracks, rust, and deformities, helping to distinguish between barriers that are functional and those requiring maintenance or replacement (S. Xu et al., 2021). For instance, (Ai & Hou, 2022) describe an automated method for detecting and evaluating guardrails using LiDAR and video-log imagery. Their approach involves georeferencing guardrail locations and extracting critical information—such as terminal types, curb presence, and lateral offset—while detecting common issues like dents, missing end terminals, and bolt damage. The system demonstrated high accuracy and efficiency in a Massachusetts pilot project, highlighting its potential for large-scale deployment in supporting Manual for Assessing Safety Hardware (MASH) compliance and asset management.

Recent advancements in guardrail detection technology address challenges such as low segmentation accuracy and noise interference. (X. Jin et al., 2024) propose an improved Unet semantic segmentation model that incorporates mixed expansion convolution and a mixed loss function to enhance detection accuracy. This model analyzes partial mean gray values in the region of interest (ROI) and employs wavelet transform's first-order detail coefficients to detect defects and deformations in guardrails. Their approach significantly improved segmentation performance, with Miou and Dice scores increasing by 8.63% and 17.67%, respectively, compared to traditional models. Additionally, their method achieved over 85% defect detection accuracy, streamlining the detection process and enhancing overall road maintenance efficiency.

Similarly, (Kasuga et al., 2020) introduced a classification method for in-vehicle camera images aimed at supporting guardrail inspections. Their approach employs Fully Convolutional Networks (FCNs) to extract guardrail regions from complex backgrounds, including vegetation, buildings, and vehicles. The model integrates rust detection to improve accuracy in identifying corroded guardrails, successfully classifying their condition as "nothing unusual," "peeling of coating," or "dirt." This system offers a reliable automated solution for infrastructure maintenance. In addition to structural health, safety barriers must meet specific design standards, such as height, length, offset, and material type, to provide adequate protection. Computer vision systems can compare detected barriers with digital blueprints or regulatory specifications to ensure they are installed correctly and meet safety guidelines (Huang et al., 2009). Barriers that are incorrectly installed or are too short may fail to provide proper protection in the event of vehicle collisions, making design adequacy a critical factor in safety evaluations.

### 2.2.3 Utilities (e.g., Signs, Culverts, Poles)

Culverts, drainage systems, and utility poles are vulnerable to damage from environmental factors such as weather conditions and traffic loads. Computer vision models, often integrated with LiDAR, are employed to assess the structural health of these utilities by detecting issues like cracks, blockages, and signs of collapse, ensuring timely interventions (Huang et al., 2009). These systems are particularly effective at identifying potential failures in underground or hard-to-access utilities, which might not be visible through conventional inspection methods.

In addition to assessing structural health, the adequacy of these utilities is equally important. Adequacy evaluation ensures that systems such as culverts and poles are designed and positioned to meet specific functional demands. For instance, culverts must be sized appropriately to manage expected water flows, while poles must be spaced adequately to maintain structural stability (Huang et al., 2009; Xie et al., 2023). Computer vision technologies, integrated with engineering guidelines, allow for the automated analysis of these utilities, confirming whether they comply with design standards. For example, (Deng et al., 2022) developed a method using Keypoint-RCNN to automatically detect and evaluate the loosening angle of bolted joints, providing high accuracy in monitoring structural safety. Additionally, (Beleznai et al., 2023) introduced a vision-based mobile analysis system that utilizes RGB-D cameras for the semantic segmentation and 3D reconstruction of roadside guardrails. This system is capable of assessing various guardrail components such as beams, bolts, and posts, measuring structural properties like alignment and spacing. Tested in real-world field environments, the system offers a practical, automated solution for large-scale monitoring, especially in critical areas like crash mitigation. Together, these advanced computer vision systems improve the detection, evaluation, and maintenance of critical infrastructure assets, ensuring both their structural health and design adequacy.

### **2.3 Summary**

Recent advancements in computer vision and deep learning have significantly enhanced the detection and condition monitoring of transportation infrastructure assets, such as safety barriers, road markings, and utilities, offering valuable insights for UDOT's project to automate barrier condition assessments. Research demonstrates that advanced algorithms and sensor technologies, including cameras and LiDAR, can accurately identify assets and assess their physical condition across diverse environments, supporting applications in road safety, autonomous driving, and asset management. These methods enable efficient monitoring of defects and compliance with design standards, reducing manual inspection efforts and improving maintenance prioritization. For UDOT, these findings underscore the potential of computer vision models to analyze large-scale image data and produce reliable condition inventories for primary assets like guardrails and concrete barriers, while also informing exploratory applications for

secondary assets such as curbs and walls, ultimately enhancing roadway safety and infrastructure efficiency.

### **3.0 Research Methods**

This research project leverages advanced computer vision and artificial intelligence to automate condition assessments of UDOT's roadside safety assets, addressing the limitations of manual inspections. Two primary methodologies are employed: object detection using the YOLO (You Only Look Once) model to identify and locate defects in primary assets (guardrails, cable barriers, and concrete barriers), and VLMs to explore enhanced condition evaluation for both primary and secondary assets (curbs, gutters, retaining walls, and shoulder edge drop-offs). These approaches utilize Pathway Services image data to develop a computer vision model for detecting the type of defects in assets, supporting UDOT's objectives of improving roadway safety, maintenance prioritization, and infrastructure management efficiency.

#### **3.1 Object Detection**

The object detection approach employs YOLO11n, a lightweight and efficient convolutional neural network (CNN)-based model from the YOLO family, to detect and classify defects in primary assets across Utah's roadways. YOLO11n was selected due to the limited number of images containing defects (e.g., guardrail with deflection or missing or broken posts) within the images collected from Pathway Services image data. As shown in Figure 1, the YOLO11 architecture consists of a backbone with efficient feature extraction layers, a neck that aggregates multi-scale features using a Feature Pyramid Network (FPN), and a head that predicts bounding boxes and class probabilities in a single stage. The streamlined design of YOLO11n, incorporating depth-wise separable convolutions and a C3 module, reduces computational complexity while maintaining high detection speed, making it ideal for datasets with sparse defect occurrences. The model is trained on an annotated subset of Pathway Services data, augmented with techniques like mosaic and flip to enhance robustness despite the limited defect samples. Transfer learning with pre-trained weights from the COCO dataset (Common Objects in Context, a large-scale dataset containing labeled images of everyday objects) is applied to improve performance on UDOT's imagery. Model performance is evaluated using mean Average Precision (mAP) and Intersection over Union (IoU) to ensure accuracy and reliability for large-scale monitoring.

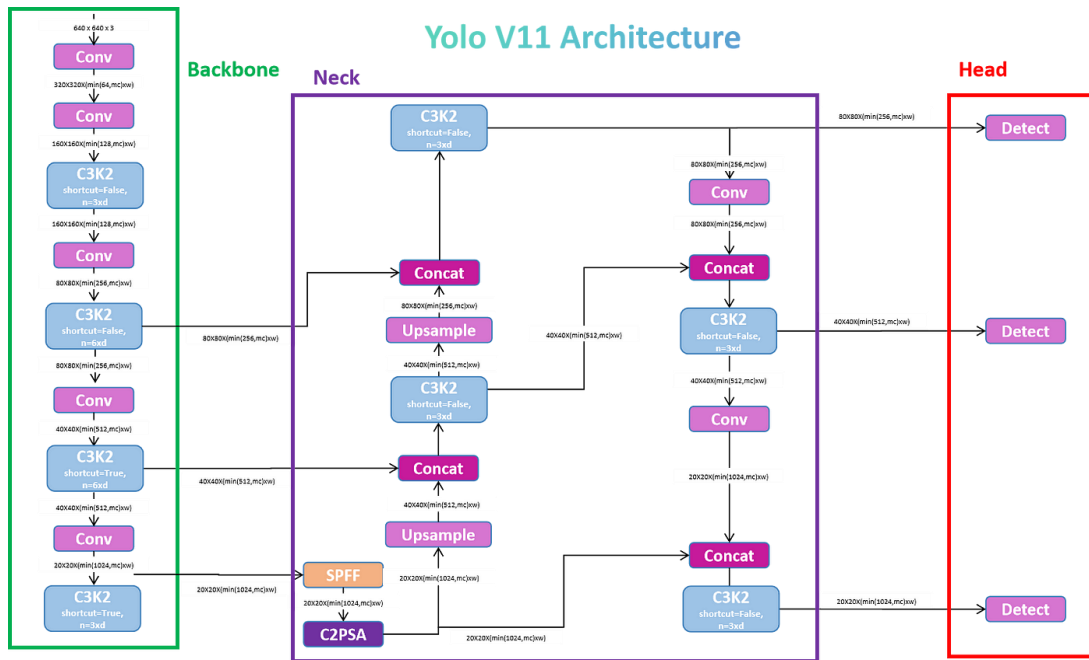


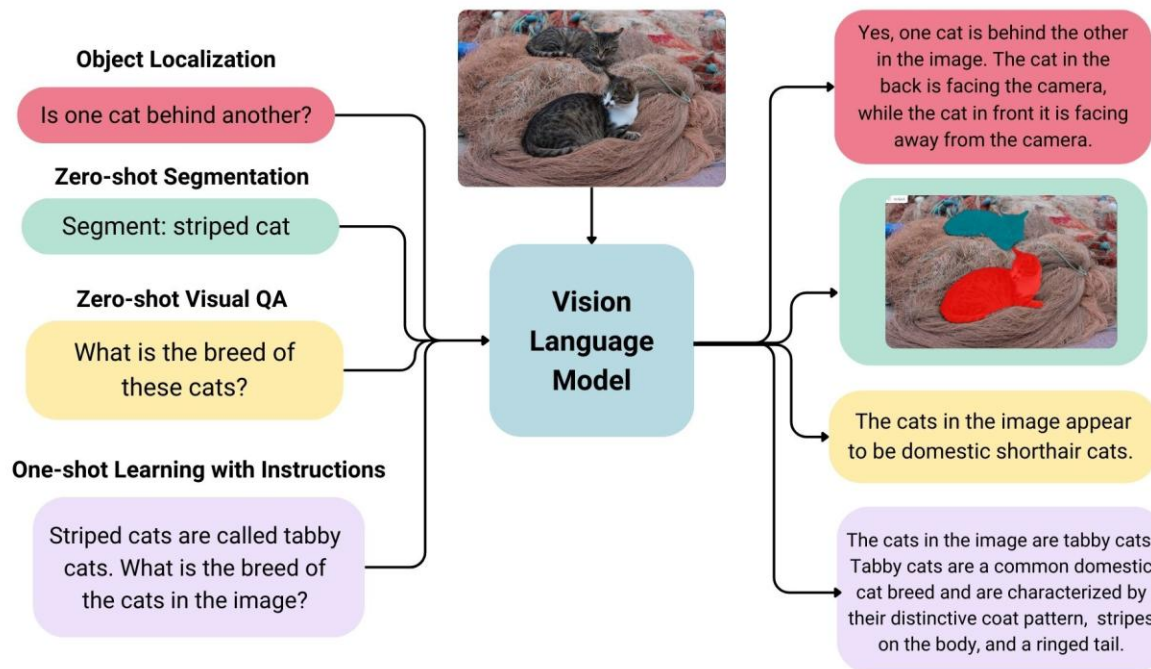
Figure 1. YOLOv11 Architecture Overview

### 3.2 Multimodal Models

Multimodal Models are an emerging class of machine learning models that integrate visual and textual data to perform a variety of computer vision tasks, including object detection, segmentation, classification, and image captioning. These models leverage the synergy between visual and linguistic information to enhance performance across diverse applications (Figure 2). A critical component of utilizing Multimodal Models for asset condition assessment is the development of tailored text prompts that guide the model to interpret asset conditions accurately. These prompts are crafted based on UDOT’s condition criteria and include descriptive phrases such as “guardrail with rust,” “concrete barrier with spalling >10%,” “missing cable barrier post,” or “curb with visible cracks.” For primary assets like guardrails, cable barriers, and concrete barriers, prompts are designed to classify defects, evaluate the condition of assets, and required maintenance activities, while also evaluating design compliance. The process involves iteratively refining prompts to improve precision, testing variations like “severely rusted guardrail” versus “guardrail with minor rust” to capture severity levels. This prompt engineering enhances the

model’s ability to distinguish subtle differences in defect states, which is essential for actionable maintenance decisions.

For secondary assets, including curbs, gutters, and retaining walls, prompt design is more flexible due to fewer standardized defect criteria. These prompts enable models to detect a broader range of issues, such as cracks, erosion, or structural shifts, by leveraging contextual understanding from text-image alignment. To evaluate performance, the pretrained models are tested on tasks such as defect classification, condition rating, and design compliance assessment. The ability to adapt prompts dynamically allows VLMs to handle the variability in secondary asset defects, offering a scalable solution for UDOT’s asset management framework. This exploratory approach aims to complement the YOLO11n object detection results, providing a richer, context-aware assessment of Utah’s transportation infrastructure.



**Figure 2. Vision Language Models Applications**

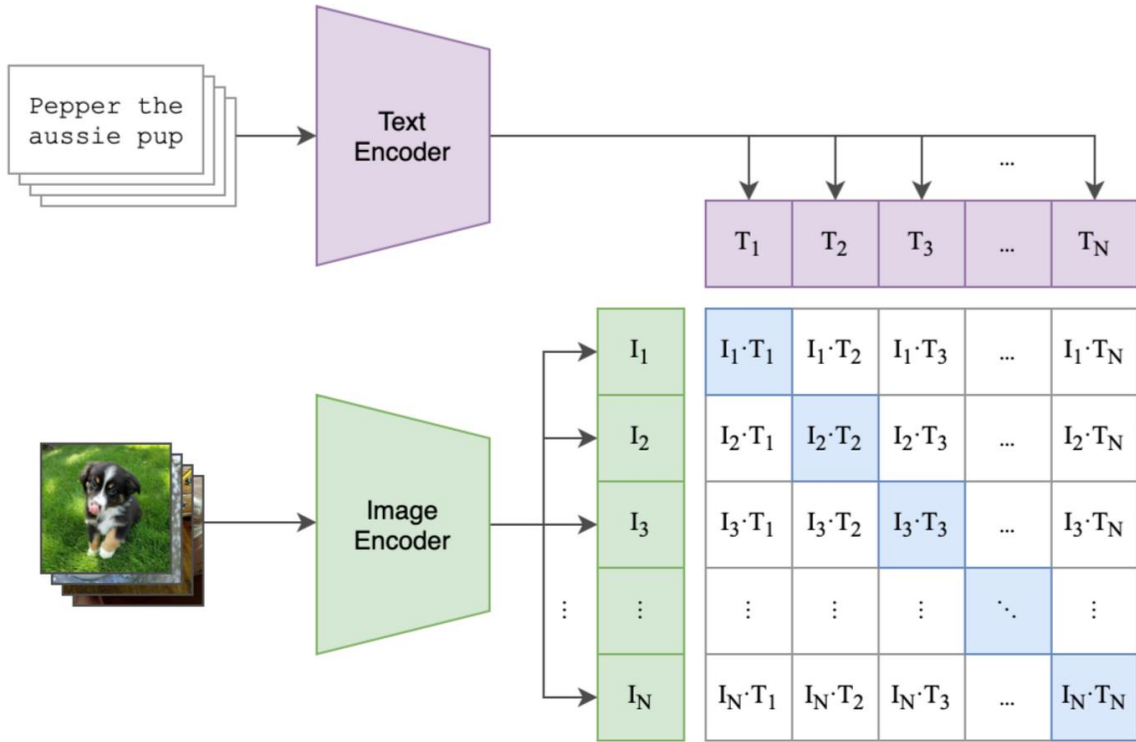
This section outlines the integration of advanced multimodal AI techniques to enhance the condition assessment process. By leveraging models that can process and analyze both visual and textual data, we achieved a comprehensive understanding of asset conditions, enabling more accurate defect identification and contextual analysis. This section details the specific approaches

employed, including zero-shot image classification and image-text-to-text models, each utilizing state-of-the-art models tailored to the assessment objectives.

### 3.2.1 Zero-Shot Image Classification

Zero-shot image classification was employed to categorize asset images into defect classes without requiring extensive labeled datasets, enabling rapid and flexible defect identification across diverse asset types. This approach leverages pretrained models that align visual and textual representations, allowing classification based on descriptive text prompts. By matching images to defect descriptions, we efficiently identified issues such as cracks, corrosion, or structural deformations, even for rare or previously unseen defects.

Several algorithms support zero-shot image classification, including CLIP (Contrastive Language-Image Pretraining) by OpenAI or Siglip by Google. These models differ in their training datasets and architecture but share the ability to perform classification without task-specific fine-tuning. In this research, we utilized the CLIP model due to its robust performance in aligning image and text embeddings, enabling accurate classification of asset defects by matching images to textual descriptions like "crack" or "corrosion." Its flexibility and generalization across asset types made it ideal for our assessment needs. CLIP employs a dual-encoder architecture comprising an image encoder (such as a Convolutional Neural Network like ResNet or a Vision Transformer) and a text encoder (typically a transformer-based model like BERT) (Figure 3). Both encoders are trained on large-scale image-text datasets, such as LAION-400M, to map visual features and textual descriptions into a shared embedding space. During inference, CLIP computes similarity scores between image embeddings and text prompts, enabling zero-shot learning where the model can perform tasks without extensive labeled data. This approach is particularly valuable for organizations like UDOT, which manages diverse asset types and variable defect conditions in data collected by services like Pathway Services.



**Figure 3. VLM Architecture Overview**

### 3.2.2 Image-Text-to-Text Models

Image-text-to-text models were applied to generate detailed textual descriptions and structured reports from asset images, automating documentation and enhancing stakeholder communication. These models process visual inputs alongside textual prompts to produce coherent narratives or formatted outputs, reducing manual effort in report generation. This approach was particularly valuable for translating complex visual defect data into actionable insights. Notable algorithms for image-text-to-text tasks include Llava (Large Language and Vision Assistant), Qwen, BLIP-2 (Bootstrapping Language-Image Pre-Training), and Flamingo. These models vary in their architecture and training, with some optimized for descriptive tasks and others for structured output generation. In this research, only the green-highlighted versions from the model comparison table were tested due to limited computational resources (RTX 4070 ~12GB GPU memory). Specifically, **Llava 1.5 7B**, **Qwen 2.5-VL 7B**, **Llama 3.2 11B**, and **Gemma 3 4B** were selected, as they provided a practical balance between efficiency and accuracy for generating detailed defect descriptions and structured reports. Although newer or larger versions of these

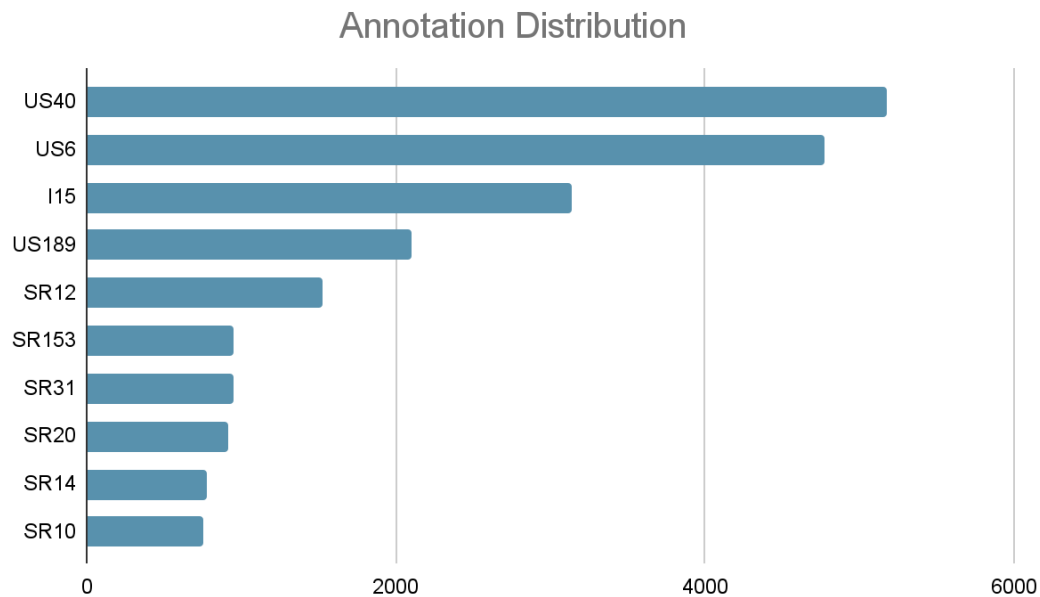
models exist, resource constraints necessitated this focus to ensure feasibility within the available infrastructure.

1. Llava 1.5: Llava 1.5 stood out for its ability to generate rich, human-like narratives from image-text pairs. Its architecture combines large-scale vision-language pretraining with instruction tuning, making it particularly effective at converting visual asset data into descriptive text that captured both technical detail and contextual nuance. This model was especially valuable for producing defect narratives that were easy for stakeholders to interpret without needing specialized domain knowledge.
2. Qwen 2.5-VL: Qwen 2.5-VL was primarily utilized for structured report generation, where precision and clarity were prioritized over expansive descriptions. Its strength lay in producing formatted, concise outputs that translated complex visual cues into tabular summaries or bullet-pointed insights. The model's balanced size enabled efficient deployment while maintaining a strong alignment between visual input and the structured textual output required for documentation.
3. Llama 3.2: The Llama 3.2 model brought a notable advantage in text generation quality, especially when complex reasoning over visual and textual information was necessary. It performed well in synthesizing detailed reports that combined multiple image inputs with textual prompts, providing logical consistency and completeness. Although it required careful resource management given its size, its output quality justified the computational cost in tasks that demanded accuracy and depth.
4. Gemma 3: Gemma 3 was selected for its versatility in handling both narrative and structured outputs at a reasonable computational footprint. While not the largest model tested, it offered a good balance of descriptive capability and report structuring, making it a reliable option for scenarios where both detailed defect description and formal documentation were needed. Its efficiency allowed it to process batches of images without significant performance trade-offs.

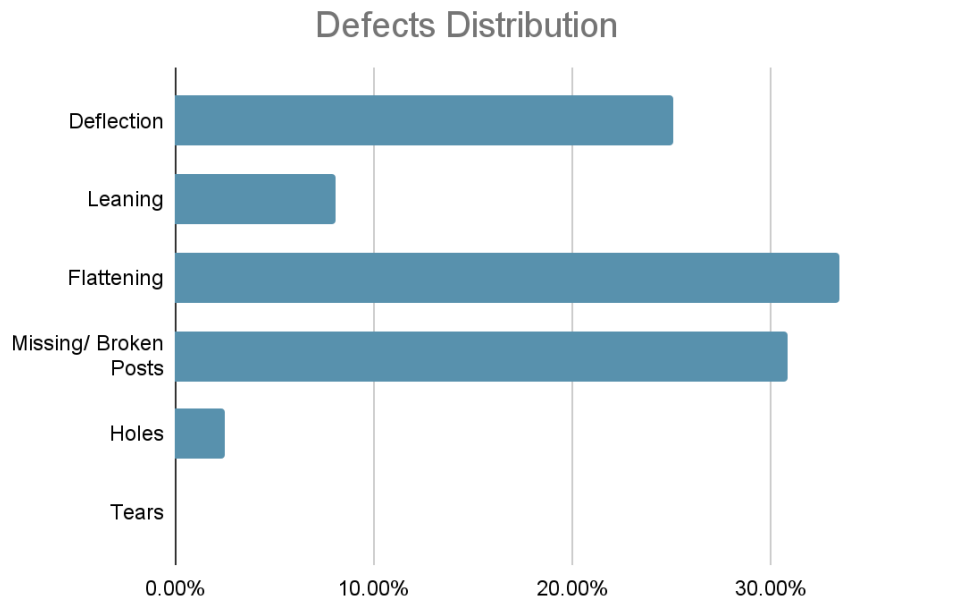
## 4.0 DATA COLLECTION

### 4.1 Overview

To develop a robust computer vision model for guardrail defect detection, a comprehensive dataset was curated from the entire state roadway network in Utah, resulting in approximately 610 miles of Utah roadways with guardrails. This data was collected using high-resolution imaging systems provided by Pathway Services, ensuring consistent visual quality across diverse environmental conditions. From this extensive collection over 100 miles of roadway imagery, around 21,000 images, were annotated. Each image was reviewed to identify and label instances of guardrail defects, adhering to standardized criteria to maintain annotation consistency. Within the annotated dataset, 470 images were identified as containing one or more defects, representing approximately 2.2% of the total annotated images.



**Figure 4. Annotated Images within Each Road**



**Figure 5. Distribution of Defects in Annotated Images**

## 4.2 Guardrail Annotation

The defect instances were categorized into six distinct classes, each corresponding to specific types of damage or anomalies observed in guardrail structures:

### 1. Deflection:

- **Good Condition:** Less than 9 inches of lateral deflection over a 25-ft length of rail.
- **Poor Condition:** More than 9 inches of lateral deflection anywhere over a 25-ft length of rail.

### 2. Leaning:

- **Good Condition:** Posts may be bent or leaning but are not broken, and guardrail sections are aligned with only slight sagging within UDOT standards for height tolerance.
- **Poor Condition:** Posts are leaning, bent, or broken, contributing to overall guardrail sagging or height below the current standard.

### 3. Flattening:

- **Good Condition:** Rail cross-section height is between 9 inches and 17 inches for the individual rail panel.
- **Poor Condition:**
  - Rail cross-section height is more than 17 inches (e.g., if the rail is flattened).
  - Rail cross-section height is less than 9 inches (e.g., due to a dent to the top edge).

### 4. Missing/Broken Posts:

- **Good Condition:** No broken posts exist.
- **Poor Condition:**
  - One or more posts are missing, cracked across the grain, broken, rotted, or torn.
  - Missing post(s) are within 62.5 ft from the end of the guardrail run, measured from the end treatment impact head or post 1 of an anchor.

### 5. Holes:

- **Good Condition:** 1–2 non-manufactured holes with a height less than 1 inch in a 12.5-ft length of rail.
- **Poor Condition:**
  - More than 2 holes with a height less than 1 inch on a 12.5-ft length of rail.
  - Any holes with a height greater than 1 inch.
  - Any hole that intersects either the top or bottom edge of the rail.

### 6. Tears:

- **Good Condition:** No vertical or horizontal tears in the rail.
- **Poor Condition:**
  - **Horizontal Tears:** Horizontal (longitudinal) tears greater than 12 inches long or greater than 0.5 inch wide. (Note: For horizontal tears less than 12 inches in length or less than 0.5 inches in height, the holes guideline applies.)
  - **Vertical Tears:** Any vertical (transverse) tear in the rail.

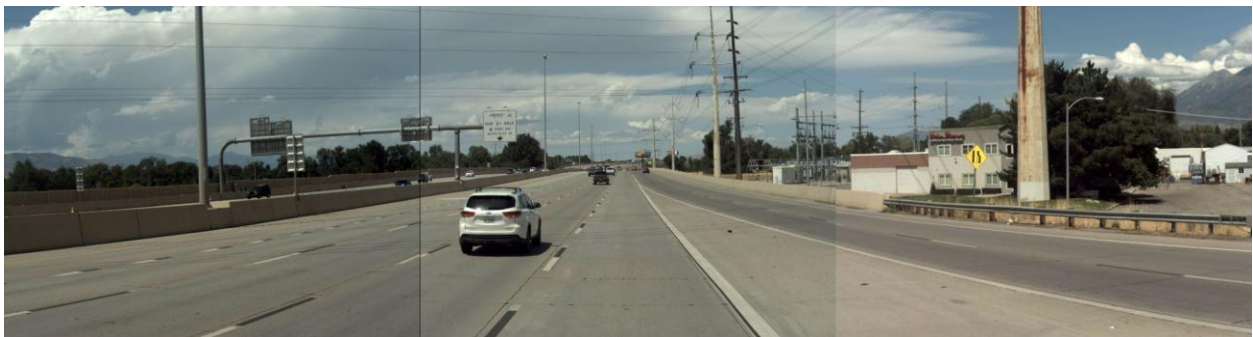


**Figure 6. Example of an Annotated Image with a Missing/Broken Post**



**Figure 7. Example of an Annotated Image with a Flattening**

Some images are hard to label because the camera angle is not good, or parts of the roadside barriers and signs are missing or hard to see. When important details are not clear, it makes it difficult for both people and AI to mark or detect the objects correctly. The image below shows one of these challenging cases where the angle and missing details make the annotation task harder.



**Figure 8. Example of a Challenging Image for Annotation**

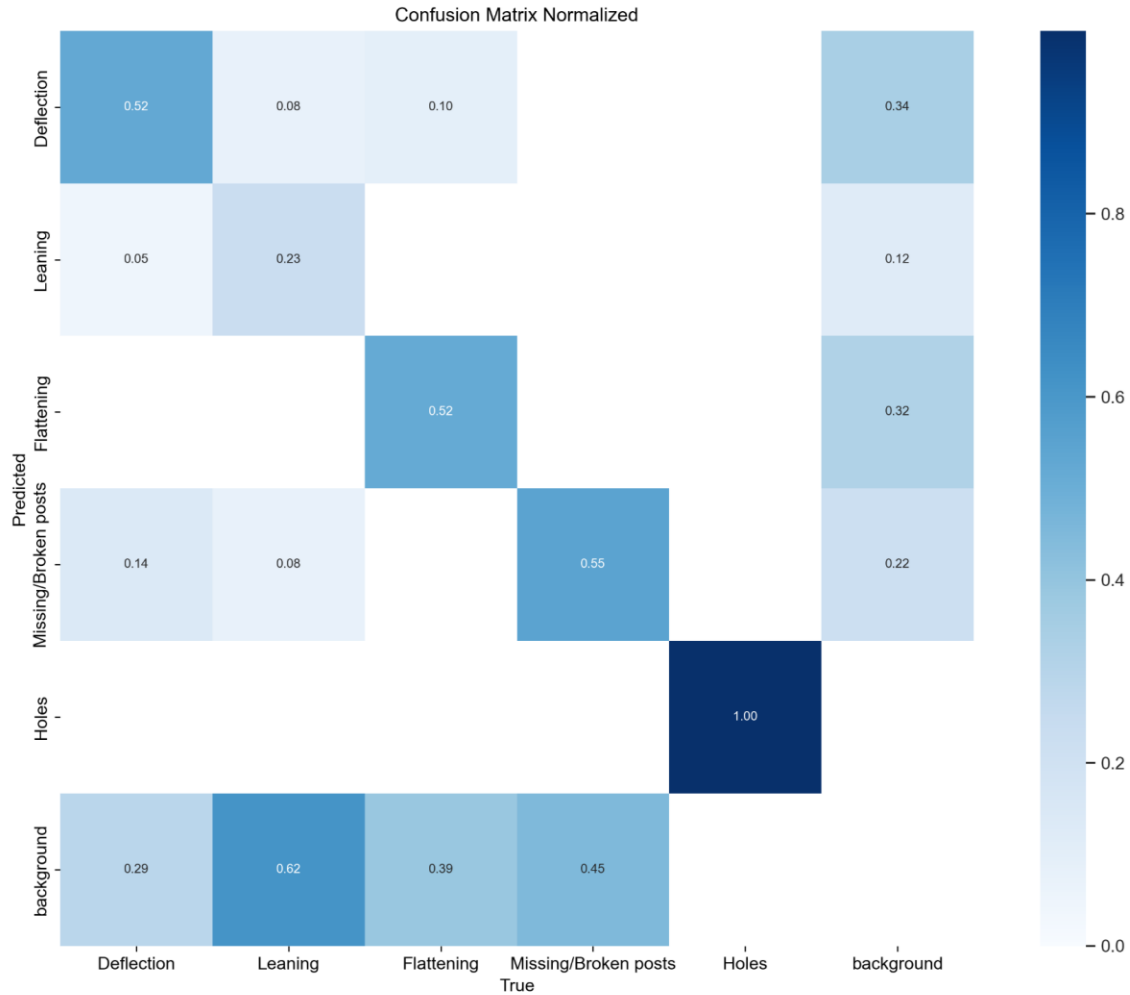
## **5.0 MODEL EVALUATION**

This section evaluates the performance of the AI models developed for UDOT’s barrier condition assessment, focusing on their ability to detect and classify defects in primary assets (guardrails, cable barriers, and concrete barriers) using Pathway Services image data. The evaluation encompasses two main approaches: object detection with the YOLO11n model, which identifies and localizes defects, and VLMs, which provide contextual condition assessments for both primary and secondary assets (curbs, gutters, retaining walls, shoulder edge drop-offs). The object detection results are presented first, assessing YOLO11n across three experimental settings—multiclass defect detection, binary classification (defect or not), and multiclass detection with data augmentation—to analyze its effectiveness. These results are followed by the VLM evaluation, which examines the model’s performance in defect classification, condition rating, and design compliance, highlighting its potential to complement object detection for a comprehensive asset management framework.

### **5.1 Object Detection Results**

#### **5.1.1 Multiclass Defect Detection**

The first evaluation setting focuses on multiclass defect detection, where YOLO11n is trained to identify and classify multiple defect types in guardrails, including deflection or missing or broken posts. This setting aims to assess the model’s ability to differentiate between various defect categories simultaneously, providing a detailed defect map for maintenance prioritization. The model was tested on a validation subset of Pathway Services data, consisting of annotated images capturing diverse defect scenarios across Utah’s roadways. Performance is measured with a mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.5 (mAP@50) of 52% and an mAP across IoU thresholds from 0.5 to 0.95 (mAP@50:95) of 25%, where mAP@50 reflects precision at a moderate overlap threshold and mAP@50:95 indicates robustness across stricter alignment criteria. A confusion matrix further evaluates classification accuracy across defect classes.

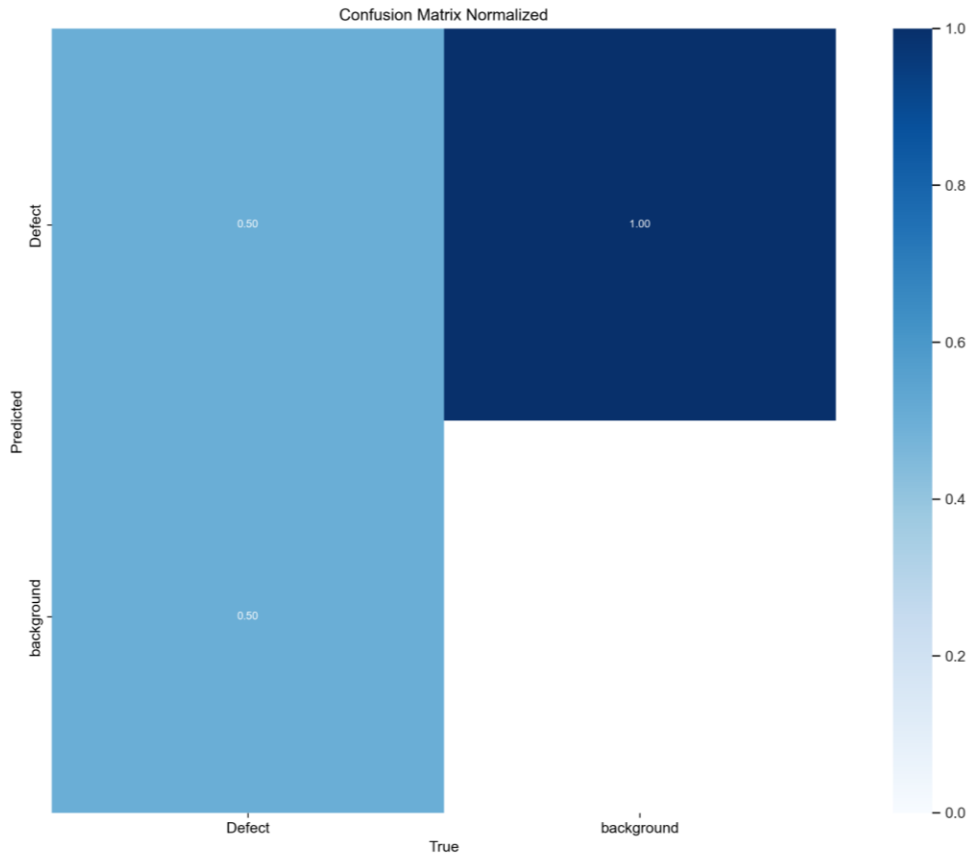


**Figure 9. Confusion Matrix for Multiclass Classification**

5.1.2 Binary Classification (Defect or Not)

The second setting simplifies the task to binary classification, where YOLO11n is trained to distinguish between images containing any defect (e.g., spalling, deflection, missing components) and those without defects in primary assets. This approach is designed to assess the model’s capability to perform a high-level screening, identifying assets requiring further inspection. The binary classification setting reduces complexity, potentially improving detection reliability given the limited number of defect images in the dataset. The model was evaluated on the same validation subset as the multiclass setting, focusing on its ability to accurately flag

defective assets for UDOT’s asset management system. Performance metrics show an  $mAP@50$  of 41% and an  $mAP@50:95$  of 15%, where  $mAP@50$  measures precision at a 0.5 IoU threshold and  $mAP@50:95$  assesses consistency across a range of thresholds from 0.5 to 0.95. A confusion matrix provides insight into true positives, false positives, true negatives, and false negatives.



**Figure 10. Confusion Matrix for Binary Classification**

### 5.1.3 Multiclass with Augmentation

The third setting extends the multiclass defect detection by incorporating data augmentation techniques to address the limited number of defect images in the Pathway Services dataset, enhancing YOLO11n’s robustness in detecting defects. Augmentation includes mosaic augmentation, which combines multiple images into a single training sample to simulate complex scenes, and flip augmentation, which mirrors images to increase data diversity, both applied to

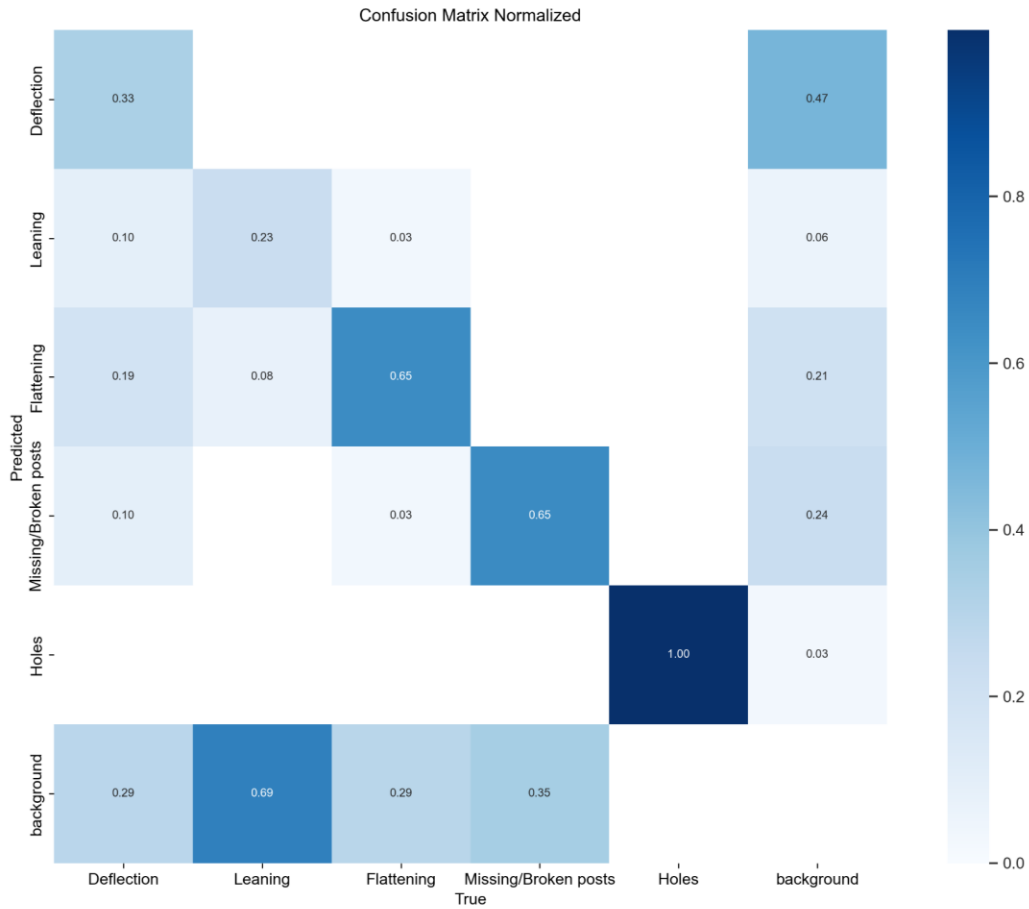
improve generalization across Utah's diverse roadway environments. Sample augmentations are illustrated in Figure 11, showing an effect applied to an image, and Figure 12, depicting a flipped image with modified contrast values, demonstrating the visual impact of these techniques. This setting aims to ensure reliable defect detection for georeferenced condition inventories, with the model evaluated on the same validation subset. Performance is assessed via an  $mAP@50$  of 63% and an  $mAP@50:95$  of 32%, where  $mAP@50$  indicates precision at a 0.5 IoU threshold and  $mAP@50:95$  evaluates performance across IoU thresholds from 0.5 to 0.95. A confusion matrix analyzes classification accuracy across defect types, highlighting the effectiveness of augmentation.



**Figure 11. Example of Image Augmentation, Up: Actual Image, Down: Augmented Image**



**Figure 12. Example of Image Augmentation, Up: Actual Image, Down: Augmented Image**



**Figure 13. Confusion Matrix for Multiclass Classification with Augmentation**

## 5.2 Multimodal Deployment

The asset assessment report presents a detailed analysis of infrastructure conditions using advanced artificial intelligence techniques. This study focuses on evaluating roadside barriers and other critical assets through multimodal model applications. The following case studies illustrate the practical application of these techniques, providing insights into defect detection and condition assessment to support maintenance and safety improvements.

## 5.2.1 CLIP

### CASE #1:



**Figure 14. CLIP CASE #1**

This case study evaluates the condition of a roadside barrier using CLIP for zero-shot image classification. The assessment analyzed an image of a guardrail and its posts against predefined prompts to determine their condition and material type. The methodology leveraged CLIP's ability to align visual and textual data for automated defect identification.

### **Prompts**

- Guardrail Condition Prompts:
  - "an intact guardrail," "a guardrail in good condition," "an undamaged guardrail"
  - "a damaged guardrail," "a guardrail with damage," "a broken guardrail"
  - "a flattened guardrail with flattening," "a flattened guardrail"
  - "a guardrail that is leaning," "a leaning guardrail," "a guardrail leaning to one side"
  - "a guardrail with holes," "a guardrail with tears," "a guardrail with holes and tears"
  - "a guardrail with missing or broken posts," "a guardrail with broken posts," "a guardrail with missing posts"

- "a guardrail with deflection," "a deflected guardrail," "a guardrail with a bend or deflection"
- Post Health Prompts:
  - "posts all intact," "no damaged posts," "posts in good condition"
  - "at least one post is damaged," "some posts are broken," "posts show signs of damage"
- Material Type Prompts:
  - "wooden posts," "posts made of wood"
  - "metal posts," "concrete posts," "posts not made of wood"

## **Responses**

The CLIP model processed the image and provided the following classification:

- Guardrail: intact
- Posts: all intact
- Material: non-wooden

This analysis highlights CLIP's effectiveness in delivering accurate condition assessments with high confidence, aligning image content with textual prompts seamlessly.

## **CASE #2:**



**Figure 15. CLIP CASE #2**

The assessment focused on identifying damage and material characteristics of a guardrail and its posts through a detailed analysis of an image, utilizing predefined textual prompts to align visual data with condition categories.

### **Prompts**

- Guardrail Condition Prompts:
  - "an intact guardrail," "a guardrail in good condition," "an undamaged guardrail"
  - "a damaged guardrail," "a guardrail with damage," "a broken guardrail"
  - "a flattened guardrail with flattening," "a flattened guardrail"
  - "a guardrail that is leaning," "a leaning guardrail," "a guardrail leaning to one side"
  - "a guardrail with holes," "a guardrail with tears," "a guardrail with holes and tears"
  - "a guardrail with missing or broken posts," "a guardrail with broken posts," "a guardrail with missing posts"
  - "a guardrail with deflection," "a deflected guardrail," "a guardrail with a bend or deflection"
- Post Health Prompts:
  - "posts all intact," "no damaged posts," "posts in good condition"

- "at least one post is damaged," "some posts are broken," "posts show signs of damage"
- Material Type Prompts:
  - "wooden posts," "posts made of wood"
  - "metal posts," "concrete posts," "posts not made of wood"

## Responses

The CLIP model processed the image and provided the following classification:

- Guardrail: damaged
- Type of Damage: deflection
- Posts: at least one damaged
- Material: wooden

This analysis underscores CLIP's capability to accurately detect and categorize damage types, such as deflection, with high probability scores.

### CASE #3:



**Figure 16. CLIP CASE #3**

The analysis focused on identifying damage and material properties of a guardrail and its posts by evaluating an image against a set of predefined textual prompts, showcasing the model's ability to detect specific damage types.

## Prompts

- Guardrail Condition Prompts:
  - "an intact guardrail," "a guardrail in good condition," "an undamaged guardrail"
  - "a damaged guardrail," "a guardrail with damage," "a broken guardrail"
  - "a flattened guardrail with flattening," "a flattened guardrail"
  - "a guardrail that is leaning," "a leaning guardrail," "a guardrail leaning to one side"
  - "a guardrail with holes," "a guardrail with tears," "a guardrail with holes and tears"
  - "a guardrail with missing or broken posts," "a guardrail with broken posts," "a guardrail with missing posts"
  - "a guardrail with deflection," "a deflected guardrail," "a guardrail with a bend or deflection"
- Post Health Prompts:
  - "posts all intact," "no damaged posts," "posts in good condition"
  - "at least one post is damaged," "some posts are broken," "posts show signs of damage"
- Material Type Prompts:
  - "wooden posts," "posts made of wood"
  - "metal posts," "concrete posts," "posts not made of wood"

## Responses

The CLIP model processed the image and provided the following classification:

- Guardrail: damaged
- Type of Damage: flattening
- Posts: at least one damaged
- Material: non-wooden

This assessment highlights CLIP's precision in identifying flattening damage, while the algorithm could not detect the material of posts correctly.

### 5.2.2 Llava 1.6

Figure 17 shows the implementation of Llava using Python and the Transformers library.

```
from transformers import pipeline

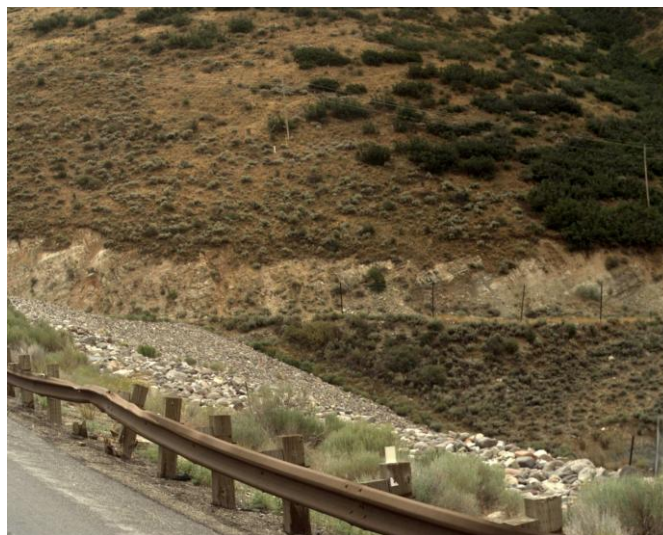
pipe = pipeline("image-text-to-text", model="llava-hf/llava-1.5-7b-hf")

messages = [
    {
        "role": "user",
        "content": [
            {"type": "image", "url": "Cable\\SR201_9_138_Roadway.jpg"},
            {"type": "text", "text": "Is the cable barrier in good condition? Is there any sign of deflection in the posts? Are the cables in tension and tight?"},
        ],
    }
]

out = pipe(text=messages, max_new_tokens=200)
print(out[0]['generated_text'][1]['content'])
```

**Figure 17. Llava Implementation in Python**

### CASE #1:



**Figure 18. Llava CASE #1**

### **Prompt**

- "Is the guardrail in good condition? Is there any defect or deflection?"

### **Response**

The guardrail in the image appears to be in good condition, with no visible defects or deflections. It is a crucial safety feature designed to protect drivers and passengers from potential accidents or falls, especially when traveling on winding roads or near steep cliffs. The well-maintained guardrail ensures that it can effectively serve its purpose and provide the necessary support and protection.

## Evaluation

Incorrect Detection: The model did not detect any deflection or missing post in the image.

## CASE #2:



**Figure 19. Llava CASE #2**

## Prompt

- "Is the concrete barrier in good condition? Is there any defect or spalling?"

## Response

The concrete barrier appears to be in poor condition, as it has a crack in it. This could potentially compromise its structural integrity and may require repair or replacement to ensure safety and prevent further deterioration.

## Evaluation

Somewhat Correct Detection: Although there are some cracks in the barrier, it is not in poor condition.

### **CASE #3:**



**Figure 20. Llava CASE #3**

#### **Prompt**

- "Is the cable barrier in good condition? Is there any sign of deflection in the posts? Are the cables in tension and tight?"

#### **Response**

The cable barrier appears to be in good condition, with no visible signs of deflection in the posts. The cables are in tension and tight, ensuring that the barrier remains stable and effective in preventing vehicles from veering off the road. The presence of multiple vehicles, including cars and trucks on the highway, indicates that the barrier is functioning as intended to maintain safety.

#### **Evaluation**

Incorrect Detection: The model did not detect the missing posts or lack of tension.

### 5.2.3 QWEN 2.5

Figure 21 shows the implementation of QWEN using Python and the Transformers library.

```
from transformers import Qwen2_5_VLForConditionalGeneration, AutoTokenizer, AutoProcessor
from qwen_vl_utils import process_vision_info

# default: Load the model on the available device(s)
model = Qwen2_5_VLForConditionalGeneration.from_pretrained(
    "Qwen/Qwen2.5-VL-7B-Instruct", torch_dtype="auto", device_map="auto"
)
# default processor
processor = AutoProcessor.from_pretrained("Qwen/Qwen2.5-VL-7B-Instruct")

messages = [
    {
        "role": "user",
        "content": [
            {
                "type": "image",
                "image": "curb&Gutter\\SR71_19.766_Roadway.jpg",
            },
            {
                "type": "text",
                "text": "Is the curb and gutters in good condition? Is there any sign of defects? Does the slope look reasonable?",
            },
        ],
    }
]

# Preparation for inference
text = processor.apply_chat_template(
    messages, tokenize=False, add_generation_prompt=True
)
image_inputs, video_inputs = process_vision_info(messages)
inputs = processor(
    text=[text],
    images=image_inputs,
    videos=video_inputs,
    padding=True,
    return_tensors="pt",
)
inputs = inputs.to("cuda")

# Inference: Generation of the output
generated_ids = model.generate(**inputs, max_new_tokens=256)
generated_ids_trimmed = [
    out_ids[len(in_ids) :] for in_ids, out_ids in zip(inputs.input_ids, generated_ids)
]
output_text = processor.batch_decode(
```

Figure 21. QWEN Implementation in Python

#### CASE #1:



Figure 22. QWEN CASE #1, Image of a Spalled Precast Concrete Barrier

#### Prompt

- "Are the concrete barriers in good health? Is there any sign of defect or spalling?"

## Response

The concrete barriers in the image appear to be in a state of disrepair and are not in good health. There are visible signs of deterioration, such as cracks and missing sections along the length of the barrier. The concrete seems to have spalled off in places, which is evident from the exposed aggregate and the uneven surface. Additionally, the vegetation growing around the barriers suggests that they may not be regularly maintained, contributing to their current condition. It's important to note that such barriers are typically designed to withstand certain levels of wear and tear but can degrade over time due to environmental factors like weathering and exposure to elements.

## Evaluation

Correct and Detailed Detection

### CASE #2:



**Figure 23. QWEN CASE #2, Image of a Spalled Precast Concrete Barrier**

## Prompt

- "Classify the condition of the concrete barrier into 1) Good, 2) Fair, 3) Poor. Also select the type of defect: 1) Scaling or Spalling, 2) Exposed Rebar, 3) One or Multiple Snag Points"

## Response

The concrete barrier in the image appears to be in a state of disrepair and is not in good condition. Here's the classification and the type of defect observed:

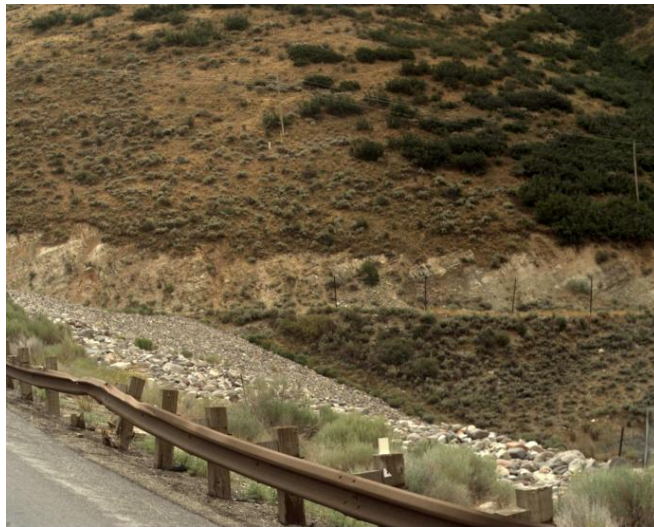
- **Condition Classification:** Poor
- **Type of Defect:** Scaling or Spalling - The concrete surface shows signs of chipping and breaking away from the main structure, which is indicative of scaling or spalling.

There are no visible signs of exposed rebar or multiple snag points in this particular section of the barrier. The damage seems localized to the surface rather than structural issues like rebar exposure or multiple weak points.

## Evaluation

Correct Detection

## CASE #3:



**Figure 24. QWEN CASE #3, Image of a Deflected Guardrail**

## Prompt

- "Is the Guardrail railing in good health? Is there any sign of defect or deflection?"

## Response

The guardrail in the image appears to be in good condition overall, but upon closer inspection, there are some signs that suggest it may not be in perfect health:

- **Deflection:** The guardrail shows noticeable deflection, particularly at the bottom where it curves. This could indicate that the guardrail has been subjected to stress, such as from heavy vehicles or environmental factors like wind.
- **Wear and Tear:** There are visible signs of wear on the metal surface of the guardrail, which might suggest exposure to the elements over time. Rust or corrosion could be present, especially if the area experiences high humidity or salt exposure (common near coastal areas).
- **Structural Integrity:** While the guardrail seems to be intact, the deflection could compromise its ability to effectively protect vehicles in an accident. It's important for guardrails to maintain their structural integrity to perform their intended function.
- **Condition of Posts:** The wooden posts supporting the guardrail appear to be in decent condition, with no obvious signs of rot or decay. However, the wood could still be susceptible to moisture damage if the area is prone to flooding or high humidity.

## **Evaluation**

Correct and Detailed Detection

## **CASE #4:**



**Figure 25. QWEN CASE #4, Image of Broken Cable Barrier Posts along a Road**

### **Prompt**

- "Are all the cable barrier posts intact? Is there any sign of defect?"

### **Response**

From the image provided, it appears that the cable barrier posts are mostly intact. However, upon closer inspection:

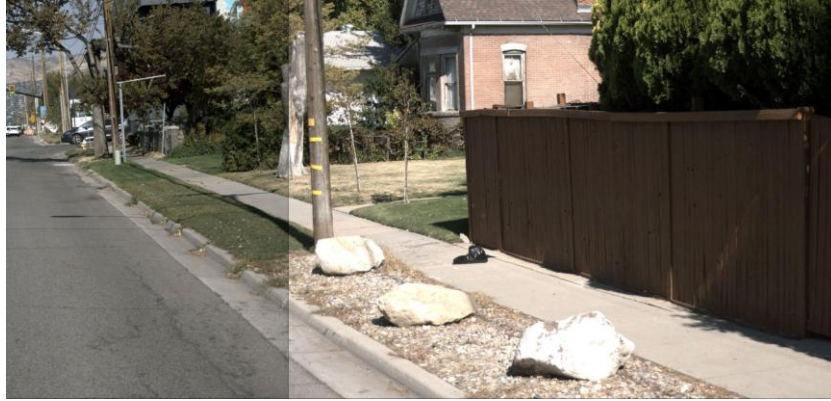
- **Top Post:** The top post of the cable barrier seems to be slightly bent or misaligned, which could indicate a defect.
- **Middle Post:** The middle post looks straight and in good condition.
- **Bottom Post:** The bottom post also appears to be straight and intact.

There is a visible sign of wear on the ground near the bottom post, but this does not seem to affect the structural integrity of the post itself. The overall condition of the cable barrier suggests it has been in place for some time, as indicated by the vegetation growing around it. If you need a more detailed assessment, it would be advisable to inspect the posts from a closer angle or with specialized equipment.

### **Evaluation**

## Correct and Detailed Detection

### CASE #5:



**Figure 26. QWEN CASE #5, Image of Curbs and Gutters in Urban Area**

### **Prompt**

- "Are the curbs and gutters in good condition? Are there any signs of defects? Does the slope look reasonable?"

### **Response**

From the images provided, here's an analysis of the curb, gutters, and slope:

- **Curb and Gutters:** The curb appears to be in good condition with no visible cracks or damage. The gutters seem to be intact and not clogged. There is no visible debris or water accumulation that would suggest blockage.
- **Slope:** The slope of the sidewalk and the area near the curb looks reasonable. It does not appear to be excessively steep or uneven, which could cause issues for pedestrians or vehicles. The transition from the street to the sidewalk seems smooth, indicating a proper slope.

Overall, both the curb and gutters appear to be in good condition without any visible defects. The slope looks reasonable and appropriate for pedestrian and vehicular traffic.

### **Evaluation**

Correct Detection

## CASE #6:



**Figure 27. QWEN CASE #6, Image of Highway Retaining Wall**

### **Prompt**

- "What kind of wall do you see in this image? Is the wall plumb? Do you see any crack or defects?"

### **Response**

The image shows a highway noise barrier wall, which is designed to reduce noise pollution from traffic for nearby residents and businesses. The wall appears to be made of a textured material that mimics the appearance of natural rock formations, likely to blend with the surrounding environment.

### **Observations:**

- **Wall Type:** It looks like a concrete or metal structure with a decorative facade that imitates rock textures.
- **Plumbness:** The wall seems to be relatively straight and vertical, but without precise measurement tools, it's difficult to determine if it is perfectly plumb. However, visually, it appears to be fairly straight.

- **Cracks or Defects:** There don't appear to be any visible cracks or significant defects on the surface of the wall. The texture looks uniform, and there are no obvious signs of damage or wear.

**Additional Notes:** The wall has a chain-link fence at the top, which might serve as a safety feature or to prevent debris from falling onto the road. The wall is supported by a concrete base, which helps in maintaining its stability and alignment. If you need more detailed information about the wall's condition, such as exact measurements or a closer inspection, it would be best to consult with a professional engineer or a construction expert.

### **Evaluation**

Incorrect Detection, but correct details

## 5.2.4 Llama 3.2

Figure 28 shows the implementation of Llama using Python and the Transformers library.

```
import requests
import torch
from PIL import Image
from transformers import MllamaForConditionalGeneration, AutoProcessor

model_id = "meta-llama/Llama-3.2-11B-Vision-Instruct"

model = MllamaForConditionalGeneration.from_pretrained(
    model_id,
    torch_dtype=torch.bfloat16,
    device_map="auto",
)
processor = AutoProcessor.from_pretrained(model_id)

url = "Cable\\SR201_9_138_Roadway.jpg"
image = Image.open(url)

messages = [
    {"role": "user", "content": [
        {"type": "image"},
        {"type": "text", "text": "Are the cable barrier posts are intact in this image? Do you see any type of damage?"}
    ]}
]
input_text = processor.apply_chat_template(messages, add_generation_prompt=True)
inputs = processor(
    image,
    input_text,
    add_special_tokens=False,
    return_tensors="pt"
).to(model.device)

output = model.generate(**inputs, max_new_tokens=300)
print(processor.decode(output[0]))
```

**Figure 28. Llama Implementation in Python**

### CASE #1:



**Figure 29. Llama CASE #1, Image of a Deflected Guardrail**

#### **Prompt:**

- “Is there any defect or deflection in the guardrail?”

#### **Response:**

There appears to be a slight deflection of the guardrail near the center of the image, but it is difficult to determine if it is a defect. The guardrail appears to be intact and securely attached to the ground, with no visible signs of damage or wear.

**Evaluation:**

Correct Detection

**CASE #2:**



**Figure 30. Llama CASE #2, Image of a Deflected Guardrail**

**Prompt:**

- “Classify the condition of this guardrail, choose from good, fair, or poor. Also select which of the following classes well define the defect in the guardrail: 1) deflection, 2) leaning, 3) flattening, 4) holes and tears, 5) missing or broken posts.”

**Response:**

To classify the condition of this guardrail, we need to analyze its components and structure. The guardrail appears to be leaning to one side, which is a significant defect. This leaning indicates that the guardrail may not be properly secured or supported, which could compromise its effectiveness in protecting vehicles from accidents. Based on the image, the following classes well define the defect in the guardrail:

**\*\*Leaning\*\***: The guardrail is noticeably tilted to one side, which suggests a lack of proper support or anchoring.

**\*\*Deflection\*\***: While there is some deflection visible, it is not as pronounced as the leaning.

**\*\*Flattening\*\***: There is no visible flattening of the guardrail.

**\*\*Holes and tears\*\***: There are no visible holes or tears in the guardrail.

**\*\*Missing or broken posts\*\***: There are no visible missing or broken posts.

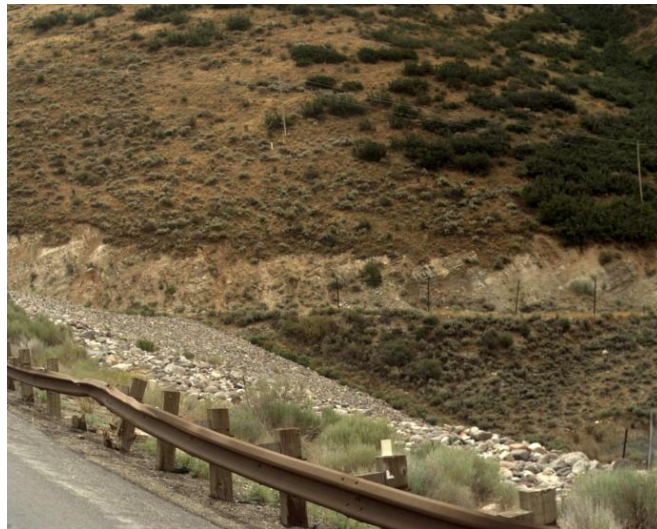
Given these observations, the condition of the guardrail can be classified as **\*\*poor\*\*** due to the significant leaning defect. The leaning is the most critical issue, as it affects the overall stability and effectiveness of the guardrail.

**\*\*Answer\*\***: Poor

### **Evaluation:**

Correct Detection

### **CASE #3:**



**Figure 31. Llama CASE #3, Image of a Deflected Guardrail**

### **Prompt:**

- “You are an expert in inspecting guardrails for defects. Analyze the following guardrail inspection criteria and identify any defects present in the guardrail. Provide a clear, concise report listing each defect detected, organized by category, and include specific details where applicable. If no defects are found in a category, state, "No defects detected" for that category. Use the provided data or observations to make your assessment.

## Guardrail Inspection Criteria:

### Deflection

Is there more than 9 inches of lateral deflection in the rail?

Is the rail leaning backwards?

Has the rail been hit and forced back?

### Flattening

Is the rail cross section more than 17 inches?

Is the face of the rail flattened?

### Posts Separated from Rail

Are all posts/blockouts attached to the rail?

### Missing/Broken Posts

Are all posts present?

Are any posts missing?

Are any posts broken?

### Blockouts

Are all blockouts present?

Are any blockouts misaligned?

Are all blockouts in good condition?

### Non-Manufactured Holes in Rail

Are there any holes in the rail larger than 1 inch? (Ignore holes for bolts)

How many holes are there in each section of rail?

### Damage at Rail Splices

Are there any missing bolts?

Have any bolts torn through the rail?

### Tears in Rail

Are there any vertical tears in the rail?

Are there any horizontal tears in the rail?

## Task:

Based on the provided guardrail data or observations (assume you have access to visual or sensor data of the guardrail), evaluate each criterion and generate a report. For each category,

list any defects found with specific details (e.g., "Lateral deflection of 12 inches detected" or "Two holes larger than 1 inch in section A"). If no defects are present in a category, explicitly state, "No defects detected." If certain data is missing or unclear, note that the assessment for that criterion cannot be completed without further information.

Output Format:

Deflection: [List defects or "No defects detected"]

Flattening: [List defects or "No defects detected"]

Posts Separated from Rail: [List defects or "No defects detected"]

Missing/Broken Posts: [List defects or "No defects detected"]

Blockouts: [List defects or "No defects detected"]

Non-Manufactured Holes in Rail: [List defects or "No defects detected"]

Damage at Rail Splices: [List defects or "No defects detected"]

Tears in Rail: [List defects or "No defects detected"]

Note: If you lack specific data to evaluate a criterion, state: "Insufficient data to assess [category/criterion]." Ensure the response is factual, precise, and avoids speculation beyond the provided criteria."

**Response:**

Deflection:

Lateral deflection of 15 inches detected.

Flattening:

No defects detected.

Posts Separated from Rail:

No defects detected.

Missing/Broken Posts:

No defects detected.

Blockouts:

Two blockouts misaligned.

Non-Manufactured Holes in Rail:

Two holes larger than 1 inch detected.

Damage at Rail Splices:

One missing bolt detected.

Tears in Rail:

No defects detected.

**\*\*Final Report:\*\***

Based on the inspection, the following defects were found:

- \* Lateral deflection of 15 inches
- \* Two misaligned blockouts
- \* Two non-manufactured holes larger than 1 inch
- \* One missing bolt at rail splice

The remaining categories showed no defects.

**Evaluation:**

Correct and Detailed Detection

**CASE #4:**



**Figure 32. Llama CASE #4, Image of Broken Cable Barrier Posts along a Road**

**Prompt:**

- “Are the cable barrier posts intact in this image? Do you see any type of damage?”

**Response:**

The cable barrier posts in this image appear to be broken or damaged. There are several posts visible on the ground, and some of them are bent or broken. This suggests that the cable barrier has been compromised and may not be functioning as intended.

**Evaluation:**

Correct Detection

**CASE #5:**



**Figure 33. Llama CASE #5, Image of Cast-in-place Concrete Barriers along a Road**

**Prompt:**

- “Are the concrete barriers in good shape? Do you see any type of damage? How about cracks? How wide do you think those are?”

**Response:**

The concrete barriers in the image appear to be in good condition, with no visible signs of significant damage or cracks. The width of the cracks is difficult to determine with certainty, but they seem to be relatively narrow, likely less than 1 inch (2.5 cm) wide.

**Evaluation:**

Correct Detection

**CASE #6:**



**Figure 34. Llama CASE #6, Mountain View and Concrete Barriers**

**Prompt:**

- “What do you see in this image?”

**Response:**

The image shows a concrete barrier running along a road that is surrounded by tall grass and evergreen trees. The road is paved and appears to be a highway or a major road. The concrete barrier is tall and has a rough texture, and it runs along the side of the road, separating it from the surrounding vegetation. The tall grass and evergreen trees are lush and green, and they appear to be well-maintained. The overall atmosphere of the image is one of serenity and tranquility, with the sound of birds chirping and the rustling of leaves in the wind.

**Evaluation:**

Correct Detection with fun details.

## CASE #7:



**Figure 35. Llama CASE #7, Noise Walls**

### **Prompt:**

- “What kind of wall do you see in this image? How is the health condition of the wall? Is there anything that we need to consider?”

### **Response:**

The wall in this image is a concrete sound barrier wall, which is designed to reduce the impact of noise from traffic on nearby residents. The health condition of the wall appears to be good, with no visible signs of damage or deterioration. However, it's important to consider the following factors:

1. **Maintenance**: Regular maintenance is crucial to ensure the wall remains effective and safe. This includes inspecting the wall for cracks, erosion, or other damage, and making repairs as needed.
2. **Weathering**: The wall may be exposed to harsh weather conditions, such as heavy rain, snow, or extreme temperatures, which can cause erosion or damage over time.
3. **Vegetation**: The presence of vegetation near the wall can lead to erosion or damage if not properly managed. Regular trimming or removal of vegetation may be necessary to prevent this.
4. **Traffic**: The wall may be subject to traffic-related damage, such as impacts from vehicles or debris. Regular inspections and repairs may be necessary to maintain the wall's integrity.

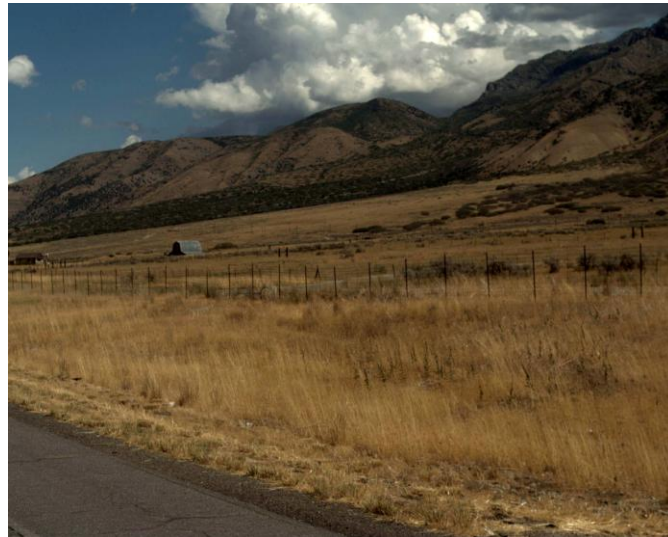
5. **Aging**: Like any structure, the wall will eventually reach the end of its lifespan. It's important to plan for replacement or refurbishment when necessary to ensure continued effectiveness and safety.

Overall, while the wall appears to be in good condition, regular maintenance and monitoring are essential to ensure its continued effectiveness and safety.

**Evaluation:**

Correct Detection

**CASE #8:**



**Figure 36. Llama CASE #8, Fences**

**Prompt:**

- “What kind of fences can you see in this image? Are they in good condition? Are all slats present? Are there any defects or deformations in them?”

**Response:**

There are two fences visible in this image, a barbed wire fence and a wire fence. The barbed wire fence is in poor condition, with broken posts and missing slats. The wire fence is also in poor condition, with some slats missing and others damaged. The barbed wire fence is located in the foreground of the image, while the wire fence is located in the background. Both fences appear to be old and worn out, with signs of rust and decay. The barbed wire fence has

several broken posts, which suggests that it may have been damaged by weather or animal activity. The wire fence has some missing slats, which could be due to wear and tear or intentional removal. Overall, both fences appear to be in need of repair or replacement.

**Evaluation:**

Incorrect Detection

**CASE #9:**



**Figure 37. Llama CASE #9, Fences**

**Prompt:**

- “What kind of fence can you see in this image? Is it in good condition? Are all slats present? Are there any defects or deformations in them?”

**Response:**

The image shows a wire fence, which is a type of fencing made from metal wire that is stretched between posts. The fence appears to be in good condition, with no visible defects or deformations in the slats. However, it's worth noting that the fence may not be as sturdy as other types of fencing, such as wooden or metal fences, and may not provide the same level of security or privacy. Additionally, the fence may not be as aesthetically pleasing as other types of fencing, and may not blend in as well with the surrounding landscape.

**Evaluation:**

Correct Detection

**CASE #10:**



**Figure 38. Llama CASE #10, Curbs and Gutters**

**Prompt:**

- “Do you see Curb or Gutter in this image? How is their condition? Are they in good condition? Is there any sign of defect or spalling?”

**Response:**

There is a Curb and Gutter visible in this image. The condition of the curb and gutter is good. The curb and gutter are well-maintained, with no visible signs of defects or spalling. The curb is straight and even, with no cracks or breaks. The gutter is clean and free of debris, with no visible signs of damage or wear. Overall, the curb and gutter appear to be in good condition and are functioning properly.

**Evaluation:**

Correct Detection

## **CASE #11:**



**Figure 39. Llama CASE #11, Precast Concrete Barriers**

### **Prompt**

- “What kind of concrete barrier do you see in this image? Precast or cast-in-place? Choose the condition from good, fair, poor. Does it comply with US standards?”

### **Response**

The concrete barrier in the image is a precast concrete barrier, specifically a Type A barrier. It is made of precast concrete and has a rectangular cross-section with a flat top and bottom. The barrier is designed to be placed on the side of the road to prevent vehicles from crossing over into oncoming traffic or other hazards.

The condition of the barrier appears to be good. It is well maintained and shows no signs of significant damage or deterioration. The concrete is smooth and free of cracks, and the barrier is securely fastened to the ground.

As for whether it complies with US standards, it is likely that this barrier meets the requirements of the American Association of State Highway and Transportation Officials (AASHTO) for precast concrete barriers. AASHTO provides guidelines for the design and construction of highway barriers, including precast concrete barriers like the one shown in the image.

\*Answer\*: Precast, Good, Yes

### **Evaluation**

## Correct Detection

### 5.2.5 Gemma 3

Figure 40 shows the implementation of GEMMA using Python and the Transformers library.

```
from transformers import AutoProcessor, Gemma3ForConditionalGeneration
import torch

def main():
    # Suppress Triton installation errors and fall back to eager execution
    # This is necessary if Triton cannot be installed on your system (e.g., Windows without specific setup)
    import torch._dynamo
    torch._dynamo.config.suppress_errors = True

    # Set float32 matmul precision for better performance if using TensorFloat32 compatible hardware
    # This is a performance optimization and not directly related to the Triton error
    torch.set_float32_matmul_precision('high')

    print("Loading model and processor...")
    try:
        model_id = "google/gemma-3-4b-it"
        # device_map="auto" attempts to use available devices like GPU, but if Triton is an issue
        # and you have a GPU, performance might be impacted without Triton optimizations.
        # If running on CPU, this might not be a major concern for the Triton error.
        model = Gemma3ForConditionalGeneration.from_pretrained(
            model_id, device_map="auto", torch_dtype=torch.bfloat16
        ).eval()
        processor = AutoProcessor.from_pretrained(model_id)
        print("Model and processor loaded.")
    except Exception as e:
        print(f"Error loading model or processor: {e}")
        return

    image_path = "Wall\\I215_1.239_Roadway.jpg"

    prompt = """
    Analyze the image of this wall. Assess its health and condition, checking for any gaps, large cracks, and broken part on the wall or if there is any other damage.
    Provide a detailed report on the barrier's condition and give recommendations for maintenance or repair.
    """

    messages = [
        {"role": "system", "content": [{"type": "text", "text": "You are an expert in civil engineering and infrastructure maintenance."}]},
        {"role": "user", "content": [
            {"type": "image", "image": image_path},
            {"type": "text", "text": prompt}
        ]}
    ]

    try:
        inputs = processor.apply_chat_template(
            messages, add_generation_prompt=True, tokenize=True,
            return_dict=True, return_tensors="pt"
        ).to(model.device, dtype=torch.bfloat16)

        with torch.inference_mode():
            output = model.generate(**inputs, max_new_tokens=600)
            decoded = processor.decode(output[0], skip_special_tokens=True)
            print("Analysis Report:", decoded)
    except Exception as e:
        print(f"Error during analysis: {e}")

if __name__ == "__main__":
    main()
```

Figure 40. GEMMA Implementation in Python

## CASE #1:



**Figure 41. GEMMA CASE #1, Image of a Deflected Guardrail**

### **Prompt**

- “You are an expert in civil engineering and infrastructure maintenance. Analyze the image of this roadside barrier. Assess its health and condition, checking for any deformations, dents, cracks, or other damage. Provide a detailed report on the barrier's condition and give recommendations for maintenance or repair.”

### **Response**

Okay, let's dive into an assessment of this roadside barrier. As a civil engineer specializing in infrastructure maintenance, I'll be looking closely at its structure, materials, and any visible signs of deterioration.

#### Image Analysis Report: Roadside Barrier Condition

**\*\*Overall Impression:** The barrier appears to be a timber-based guardrail, likely constructed some years ago, given the weathered appearance. It's situated in a roadside environment, suggesting it's likely exposed to significant environmental factors – moisture, temperature fluctuations, and potential impact from vehicles.

**\*\*Detailed Assessment:**

1. Material – Timber: The barrier is constructed of timber. Judging by the color and texture, it appears to be treated pine or a similar softwood species. The weathering is considerable, indicating significant exposure.

2. Post Condition:

Overall: The posts themselves look to be in reasonable condition overall, though there's clear evidence of weathering.

Dents & Damage: There's a noticeable dent or impact mark on the lower post near the front of the barrier (visible on the left side of the image). This suggests a past impact event – likely from a vehicle.

Rot/Splitting: Careful inspection (difficult to do just from an image) would be needed to determine the extent of rot or splitting within the posts. Areas of discoloration, soft spots, or vertical cracks would typically indicate potential structural compromise and may require closer examination.

## Evaluation

Correct Detection

### CASE #2:



**Figure 42. GEMMA CASE #2, Image of a Spalled Precast Concrete Barrier**

**Prompt**

- “You are an expert in civil engineering and infrastructure maintenance. Analyze the image of this roadside barrier. Assess its health and condition, checking for any deformations, dents, cracks, or other damage. Provide a detailed report on the barrier's condition and give recommendations for maintenance or repair.”

## **Response**

Okay, let's analyze this concrete barrier. As a civil engineer specializing in infrastructure maintenance, here's my detailed assessment based on the image provided:

### **\*\*Overall Impression:**

The barrier appears to be a substantial concrete retaining wall or channel lining, likely part of a roadway or drainage system. It's showing signs of aging and degradation, particularly associated with exposure to the elements and potentially freeze-thaw cycles if located in a colder climate. The condition is concerning and warrants prompt investigation and potentially significant maintenance.

### **\*\*Detailed Condition Assessment:**

#### **1. Surface Condition & Spalling:**

**Significant Spalling:** The most noticeable issue is the extensive spalling – the flaking and crumbling of the concrete surface. This is most prevalent on the right side of the barrier, and also evident on the left. Spalling is a symptom of concrete deterioration, often caused by freeze-thaw cycles, chloride ingress (salt penetration), and abrasion.

**Uniform Deterioration:** The spalling doesn't seem to be concentrated in one specific area, suggesting a relatively widespread problem affecting the entire barrier.

#### **2. Rebar Exposure:**

**Moderate Exposure:** There is visible exposed rebar in several locations. Notably, you can see exposed rebar at the base of the barrier and around the areas of heaviest spalling. This is a critical

concern. Exposed rebar is vulnerable to corrosion, which weakens the concrete matrix and accelerates deterioration.

### 3. Cracking:

**Fine Cracking:** There are numerous hairline cracks visible across the surface. These are likely shrinkage cracks, common in concrete as it dries and cures. However, their prevalence alongside the spalling suggests they may be connected to the larger deterioration processes.

**Lack of Major Cracks:** There are no large, open cracks that would indicate significant structural movement or distress.

### 4. Pin Slots:

**Present:** I can clearly see several pin slots running along the length of the barrier. These slots are used to secure the barrier to the underlying structure.

### 5. Structural Integrity (Visual):

**General Stability:** Based on the photo, the barrier appears largely stable. However, the spalling is compromising its protective function and weakening the structure. A thorough structural assessment, including deflection measurements and potentially non-destructive testing (NDT), would be needed to confirm its long-term stability.

### 6. Vegetation:

**Vegetation Growth:** The presence of weeds and small bushes growing around and through the barrier indicates that the barrier is not effectively preventing vegetation encroachment. This can contribute to further damage by adding weight and potentially introducing root growth.

### \*\*Recommendations for Maintenance and Repair:

**Immediate Stabilization:** Implement temporary measures (e.g., shotcrete application) to stabilize the loose concrete and prevent further material loss before the next rainfall

## **Evaluation**

Correct and Detailed Analysis

## **6.0 CONCLUSIONS**

Manual inspections of roadside safety assets, such as guardrails, cable barriers, and concrete barriers, are labor intensive, time consuming, and prone to human error, often leading to inconsistent condition assessments and delayed maintenance prioritization for UDOT. The need for automation led this project to leverage advanced AI techniques for a scalable solution for barrier condition assessment. The YOLO11n model delivered poor performance across three settings: multiclass defect detection resulted in an mAP@50 of 52% and mAP@50:95 of 25%, binary classification yielded an mAP@50 of 41% and mAP@50:95 of 15%, and multiclass with augmentation improved to an mAP@50 of 63% and mAP@50:95 of 32%. These results highlighted the limitations of conventional object detection models and motivated the research team to utilize more advanced and groundbreaking algorithms. Most of the tested multimodal algorithms showed promising results with detailed and contextually accurate responses. Among them, Gemma 3 achieved the highest accuracy but required the most time, with each image taking approximately 30 minutes on the current GPU. Llama 3.2 delivered the best performance overall, balancing both processing time and accuracy, due in part to the use of a more advanced version with 11B parameters.

### **6.1 Recommendations**

In terms of deployment, each run (single image) required approximately 4–6 minutes to complete (Llama 3.2), leading to an estimated 1.5 to 2 hours of processing time per mile (assuming 20 images in each mile taken at a tenth of a mile). However, the actual run time depended heavily on the available computational resources. For future directions:

- Focus on LLama 3.2 (or newer versions) for operational deployment, as it delivered the best balance of speed and accuracy.
- Prioritize investment in computational resources or explore cloud-based solutions to support larger-scale deployments of advanced models like Gemma 27B or Llava 4.
- Continue monitoring emerging models (e.g., Gemma 3N) for potential integration, as they offer better efficiency with lower resource demands.

- Consider using premium models such as GPT-4o for critical tasks where enhanced capabilities justify the cost.
- Maintain a clear focus on specific, high-impact use cases to ensure the best utilization of available resources.
- Developing correct prompts can significantly improve results. Any future research should first test various prompts with the selected model to determine which provides the best results.

## 6.2 Limitations

All implementations were done using Python with open-source tools, though premium models like GPT-4o offer enhanced capabilities for a fee. Advanced multimodal models, such as Gemma 27B or Llava 4, demonstrated superior performance but required significant computational resources. Lightweight versions of these models have been effectively deployed on local desktops.

1. **The field continues to evolve rapidly, with new models like Gemma 3N emerging, offering efficient performance with minimal resource demands.** This rapid evolution necessitates ongoing updates to the system to incorporate cutting-edge models, which can be challenging given limited development timelines and resources. Models like Gemma 3N promise reduced computational needs, making them viable for broader adoption, but their integration requires continuous evaluation and adaptation.
2. **Hallucination risk: Some models occasionally produced fabricated or irrelevant outputs, particularly when prompts were too complex or vague (Llama CASE #3):** Hallucination, where multimodal models generate plausible but incorrect data (e.g., inventing damage details not present in images), poses a significant risk. For instance, in complex scenarios involving multiple barrier types, models might falsely report extensive damage or nonexistent features, undermining trust in automated assessments and potentially leading to misinformed maintenance decisions. This issue was more pronounced with less constrained inputs, highlighting the need for careful prompt engineering.

3. **Prompt sensitivity: More detailed and precise prompts consistently resulted in more accurate and useful answers:** The quality of responses varied significantly with prompt design. For example, in Llama Case #9, using the singular noun "fence" in the prompt yielded a focused assessment of a single structure, while Case #8 with the plural "fences" led to a broader but less precise analysis, sometimes missing specific defects. A general prompt like "Assess the barrier" produced a vague overview, while a too complicated prompt like "Evaluate the guardrail's condition, material composition, installation date, and future maintenance needs with precise measurements" resulted in a convoluted response with hallucinated details (e.g., fabricated installation dates), illustrating a tradeoff between detail and accuracy.
4. **Computational demands: Advanced multimodal models required substantial processing time and hardware resources, which constrained large-scale deployment within the current budget.** Models like Gemma demanded high-end GPUs, with each image analysis taking up to 30 minutes, limiting the project's ability to process a wide range of images. This resource intensity restricted scalability, necessitating a focus on optimizing lightweight models or exploring cost-effective cloud solutions to balance performance and budget constraints.

## **7.0 REFERENCES**

- Ai, C., & Hou, Q. (2022). *Automated Guardrail Inventory and Condition Evaluation*. Massachusetts. Dept. of Transportation. Office of Transportation Planning.
- Balali, V., & Golparvar-Fard, M. (2015). Segmentation and recognition of roadway assets from car-mounted camera video streams using a scalable non-parametric image parsing method. *Automation in Construction*, 49, 27–39.
- Balali, V., & Golparvar-Fard, M. (2016). Evaluation of multiclass traffic sign detection and classification methods for US roadway asset inventory management. *Journal of Computing in Civil Engineering*, 30(2), 04015022.
- Beleznai, C., Göbel, K., Stefan, C., Dorninger, P., & Pusica, A. (2023). Vision-based mobile analysis of roadside guardrail structures. *Proceedings of the 2023 6th International Conference on Machine Vision and Applications*, 167–173.
- Chai, J., Zeng, H., Li, A., & Ngai, E. W. T. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, 6, 100134.
- Chian, E., Fang, W., Goh, Y. M., & Tian, J. (2021). Computer vision approaches for detecting missing barricades. *Automation in Construction*, 131, 103862.
- Deng, X., Liu, J., Gong, H., & Huang, J. (2022). Detection of loosening angle for mark bolted joints with computer vision and geometric imaging. *Automation in Construction*, 142, 104517.
- Gumaelius, N. (2022). *Guardrail detection for landmark-based localization*.
- Guo, B. H. W., Zou, Y., Fang, Y., Goh, Y. M., & Zou, P. X. W. (2021). Computer vision technologies for safety science and management in construction: A critical review and future research directions. *Safety Science*, 135, 105130.
- Huang, A. S., Moore, D., Antone, M., Olson, E., & Teller, S. (2009). Finding multiple lanes in urban road networks with vision and lidar. *Autonomous Robots*, 26, 103–122.
- Jin, H., Lan, Z., & He, X. (2021). On Highway Guardrail Detection Algorithm Based on Mask RCNN in Complex Environments. *2021 7th International Conference on Systems and Informatics (ICSAI)*, 1–6.
- Jin, X., Gao, M., Li, D., & Zhao, T. (2024). Damage detection of road domain waveform guardrail structure based on machine learning multi-module fusion. *Plos One*, 19(3), e0299116.

- Kasuga, K., Harakawa, R., & Iwahashi, M. (2020). Classification of In-Vehicle Camera Images for Guardrail Inspection Support. *2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech)*, 113–114.
- Kolar, Z., Chen, H., & Luo, X. (2018). Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. *Automation in Construction*, 89, 58–70.
- Liu, Y., Wang, Y., & Li, X. (2019). Computer vision technologies and machine learning algorithms for construction safety management: A critical review. *International Conference on Construction and Real Estate Management 2019*, 67–81.
- Luo, Z., Gao, L., Xiang, H., & Li, J. (2023). Road object detection for HD map: Full-element survey, analysis and perspectives. *ISPRS Journal of Photogrammetry and Remote Sensing*, 197, 122–144.
- Mamun, A. Al, Ping, E. P., Hossen, J., Tahabilder, A., & Jahan, B. (2022). A comprehensive review on lane marking detection using deep neural networks. *Sensors*, 22(19), 7682.
- Matsumoto, H., Mori, Y., & Masuda, H. (2021). Extraction of Guardrails from MMS Data Using Convolutional Neural Network. *International Journal of Automation Technology*, 15(3), 258–267.
- Mostafa, K., & Hegazy, T. (2021). Review of image-based analysis and applications in construction. *Automation in Construction*, 122, 103516.
- Paneru, S., & Jeelani, I. (2021). Computer vision applications in construction: Current state, opportunities & challenges. *Automation in Construction*, 132, 103940.
- Peinado, H. S., Melo, R. R. S., Santos, M. C. F., & Costa, D. B. (2023). Potential application of Deep Learning and UAS for guardrail safety inspections. *Annual Conference of the International Group for Lean Construction*, 31.
- Scharwächter, T., Schuler, M., & Franke, U. (2014). Visual guard rail detection for advanced highway assistance systems. *2014 IEEE Intelligent Vehicles Symposium Proceedings*, 900–905.
- Strain, T., Eddie Wilson, R., & Littleworth, R. (2020). Computer vision for rapid updating of the highway asset inventory. *Transportation Research Record*, 2674(9), 245–255.
- Vidal, M., Díaz-Vilariño, L., Arias, P., & Balado, J. (2020). Barrier and guardrail extraction and classification from point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 157–162.

- Xie, K., Sun, H., Dong, X., Yang, H., & Yu, H. (2023). Automating intersection marking data collection and condition assessment at scale with an artificial intelligence-powered system. *Computational Urban Science*, 3(1), 24.
- Xu, S., Wang, J., Wu, P., Shou, W., Wang, X., & Chen, M. (2021). Vision-based pavement marking detection and condition assessment—A case study. *Applied Sciences*, 11(7), 3152.
- Xu, Z., Cao, J., & Lan, Z. (2020). On Highway Guardrail Segmentation Algorithms Based on Patrol Car Mobile Video in Complex Environments. *2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS)*, 1158–1164.
- Yu, Y., Jiang, T., Li, Y., Guan, H., Li, D., Chen, L., Yu, C., Gao, L., Gao, S., & Li, J. (2022). SignHRNet: Street-level traffic signs recognition with an attentive semi-anchoring guided high-resolution network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 192, 142–160.