CRACKATHON: Outlining our methodology.

A deep dive into our systematic approach to high-performance road damage detection.

Meet the Team

CYPHERFORCE

VAISHNAVI PAWAR

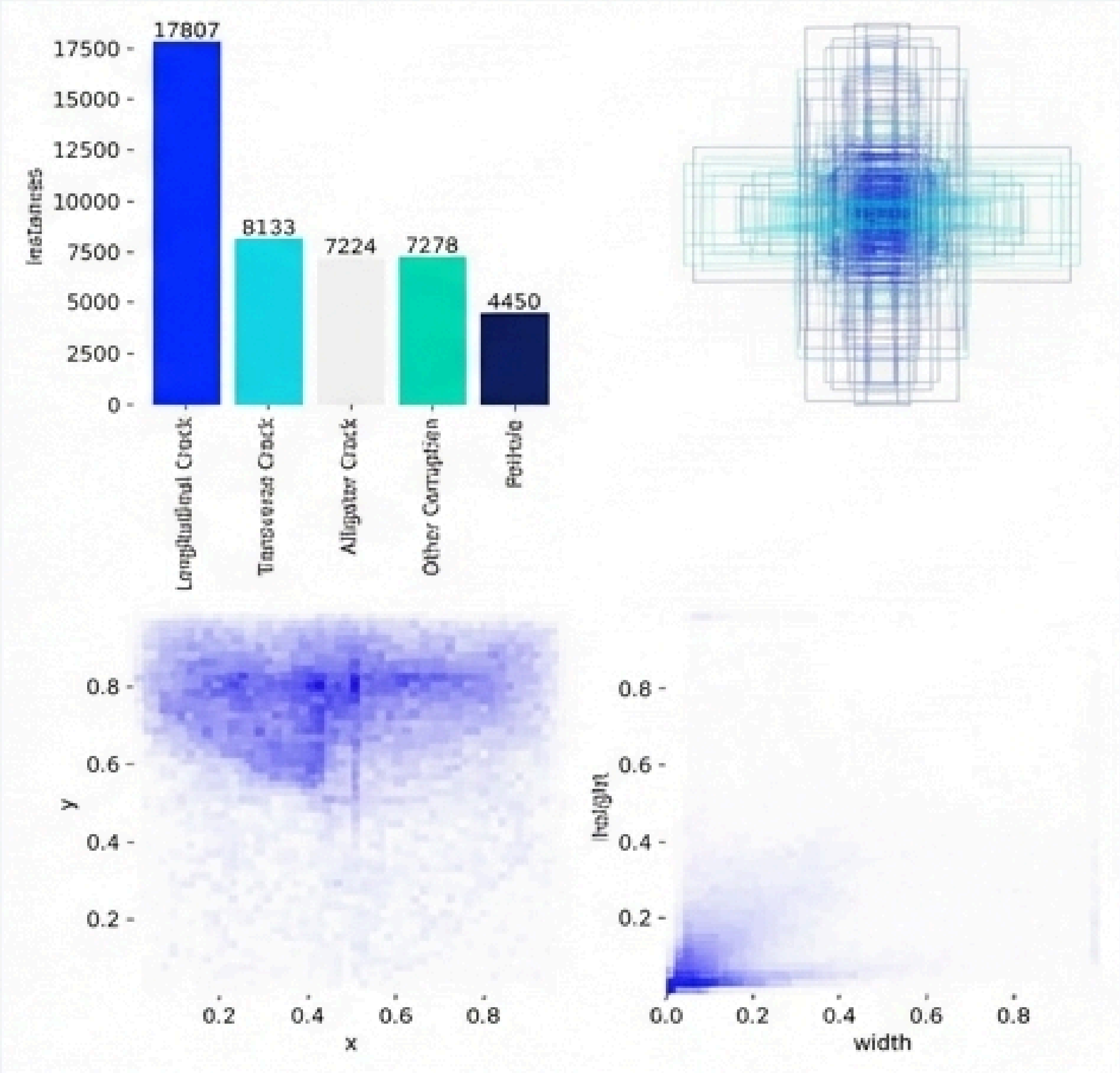
VISHWAJEET PAWAR

The Challenge: Precision & Scale in Road Damage Detection

The core problem is defined by three key difficulties:

- **Significant Class Imbalance:** As shown in the data, certain defects like "Longitudinal Crack" vastly outnumber others, creating a risk of model bias.
- **Feature Ambiguity:** Visually similar classes, such as *Longitudinal* vs. *Transverse* cracks, demand a model with high discriminative power.
- **Variable Scale:** Defects range from large potholes to fine, sub-pixel hairline cracks, requiring a multi-scale detection strategy.

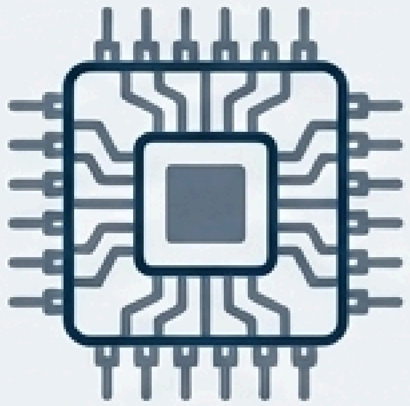
Visual Analysis: Dataset Characteristics



This visualization of our dataset's statistics highlights the core challenges. The bar chart confirms the class imbalance, while the heatmaps show the distribution of bounding box sizes and locations, which informed our model selection and training strategy.

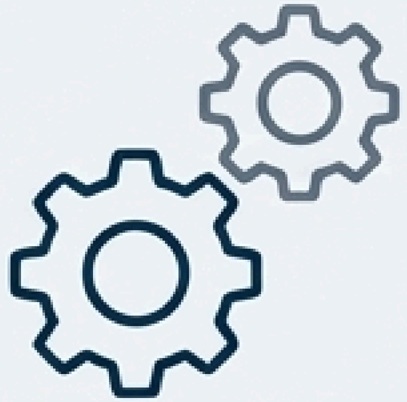
Our Solution: A Three-Pillar Engineering Framework

Our methodology is built on a foundation of three interconnected pillars, each designed to address a specific aspect of the challenge.



Pillar 1: High-Capacity Model

We selected the **YOLOv8-XLarge** architecture for its vast parameter space, enabling it to learn the subtle textural differences between challenging crack types.



Pillar 2: Resilient Training Pipeline

A multi-stage **data augmentation** strategy and robust environment handling were engineered to maximize learning and prevent overfitting.

Pillar 3: Specialized Inference Engine

We developed a custom engine using **SAHI + NMM** to achieve high-recall detection of minute features in high-resolution imagery.

A systematic approach designed for precision, recall, and robustness.

Pillar 1: Selecting a High-Capacity Architecture

Our initial analysis revealed that lightweight models lack the parameter space to distinguish between visually similar defects like *Longitudinal* and *Transverse* cracks. A larger, more capable model was essential.

Model Choice: YOLOv8-XLarge

- **Rationale:** The model's ~68 million parameters provide the necessary capacity to encode the fine-grained textural variances required for high-fidelity detection.
- **Backbone:** Utilizes a modified CSPDarknet53 with advanced C2f modules for efficient feature extraction.
- **Initialization:** We leveraged **COCO pre-trained weights** to accelerate convergence and improve the model's generalization capabilities from the start.

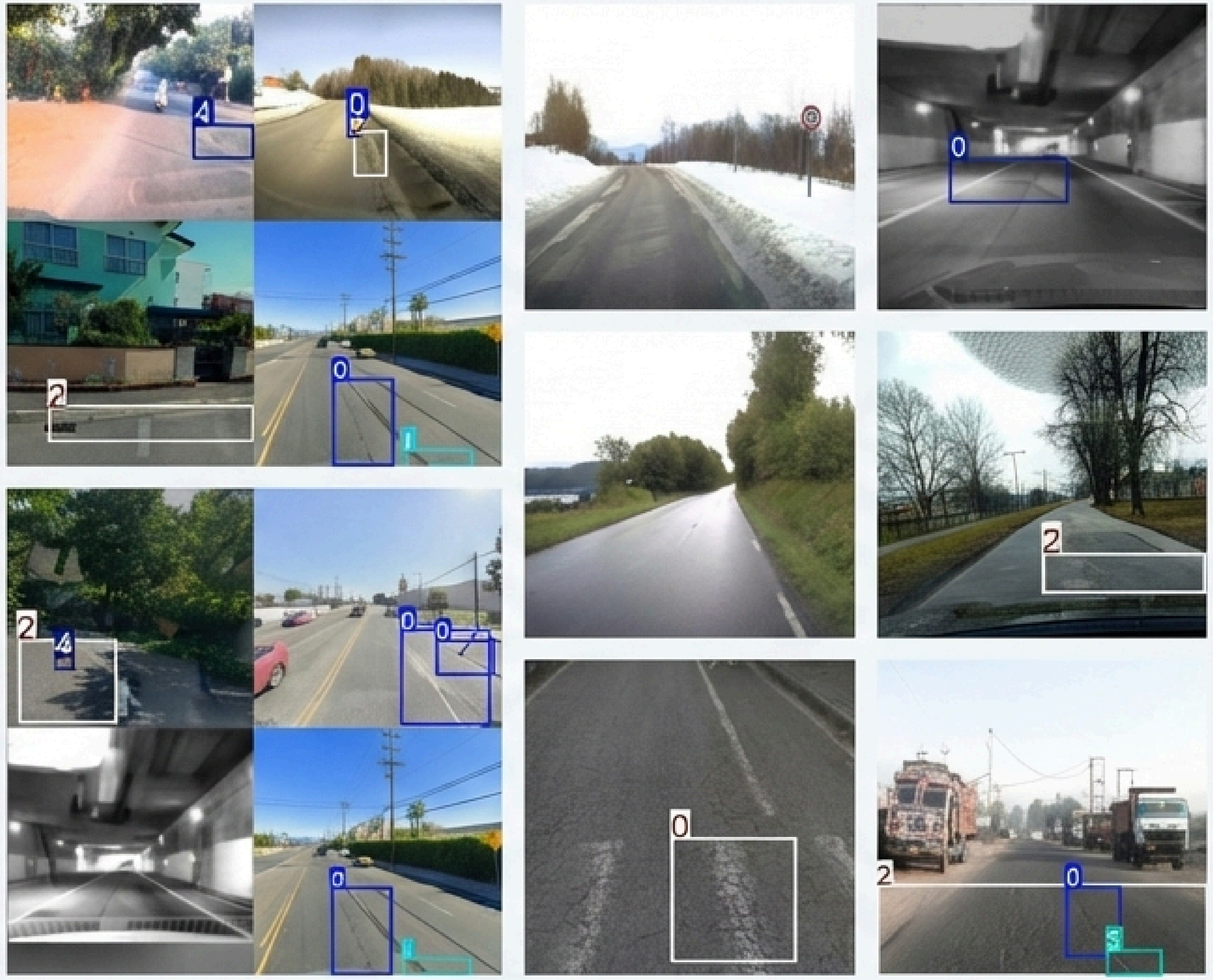
Pillar 2: Engineering a Resilient Training Pipeline

To build a model that generalizes across diverse conditions, we engineered a robust, multi-stage augmentation pipeline.

Our Augmentation Strategy:

- **1. Mosaic Augmentation (Epochs 1-40):** Stitched four images together, forcing the model to learn object detection in varied contexts and scales.
- **2. MixUp & HSV Shifts:** Blended images and adjusted color values to simulate different lighting and environmental conditions, softening decision boundaries.
- **3. Close Mosaic Optimization (Epochs 41-50):** We deactivated Mosaic for the final 10 epochs. This critical step allowed the model to fine-tune on natural image statistics, significantly reducing false positive rates.

Training Batch Examples



These samples from our training batches demonstrate the complex, augmented scenes our model learned from. The labels indicate different classes (e.g., 0: Longitudinal Crack, 4: Pothole).

Training Deep Dive: The Convergence Story

Our model’s training progress shows a clear narrative of adaptation and steady improvement, culminating in a final validation **mAP50 of 0.64**.

Key Observations:

- **Initial Volatility (Epochs 1-3):** The initial performance dip is a direct result of the aggressive **Mosaic augmentation**. The model is adapting to the synthetic, four-in-one image contexts.
- **Robust Monotonic Increase:** Following this adaptation phase, the model's performance shows a strong and consistent upward trend, confirming that our learning rate and augmentation strategy were effective.

The Inference Challenge: Detecting Hairline Cracks in HD Images

Standard inference methods fail when dealing with high-resolution road imagery.

The Problem with Downscaling:

- Resizing a large image to a model's native input size (e.g., 640x640) can completely erase fine, sub-pixel features like hairline cracks.
- Linear features like long cracks are often bisected at tile edges in naive tiling approaches, leading to fragmented, low-confidence detections.

A specialized inference engine is required to overcome these limitations.

High-resolution images contain both large-scale context and minute details. A successful inference strategy must capture both without sacrificing fidelity.

Pillar 3: Our Solution, Part 1 — Slicing Aided Hyper Inference (SAHI)

To preserve detail, we slice high-resolution images into overlapping tiles and run inference on each one.

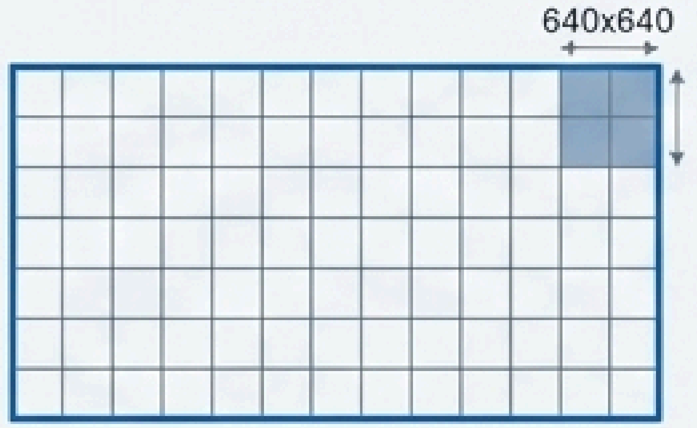
Our Optimized SAHI Process:

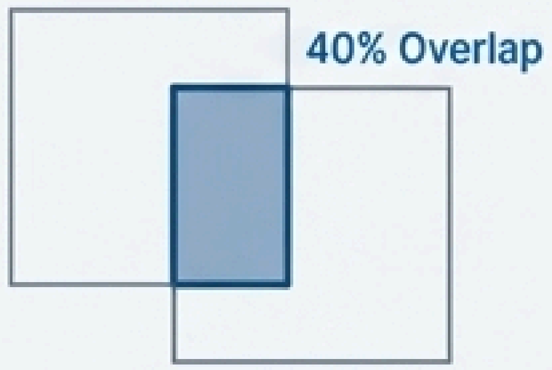
1. Slice Image

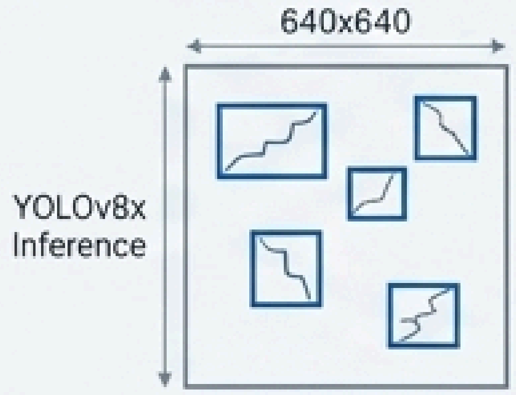
The full-resolution input is divided into 640x640 tiles, matching the model's training resolution.

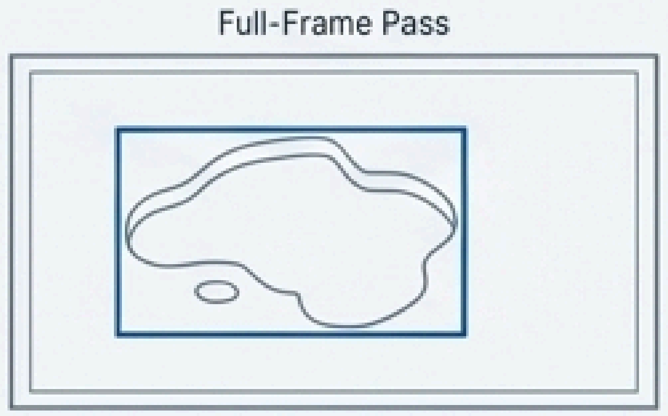
3. Run Tiled Inference

The YOLOv8x model predicts on each individual tile, allowing it to detect the smallest defects at full resolution.









1. Slice Image

The full-resolution input is divided into 640x640 tiles, matching the model’s training resolution.

2. Optimize Overlap (40%)

We use a **40% overlap ratio between tiles**. This is a critical optimization. Our analysis showed a standard 20% overlap frequently cut cracks in half. 40% ensures the entirety of most linear features is captured in at least one tile.

3. Run Tiled Inference

The YOLOv8x model predicts on each individual tile, allowing it to detect the smallest defects at full resolution.

4. Run Full-Frame Pass

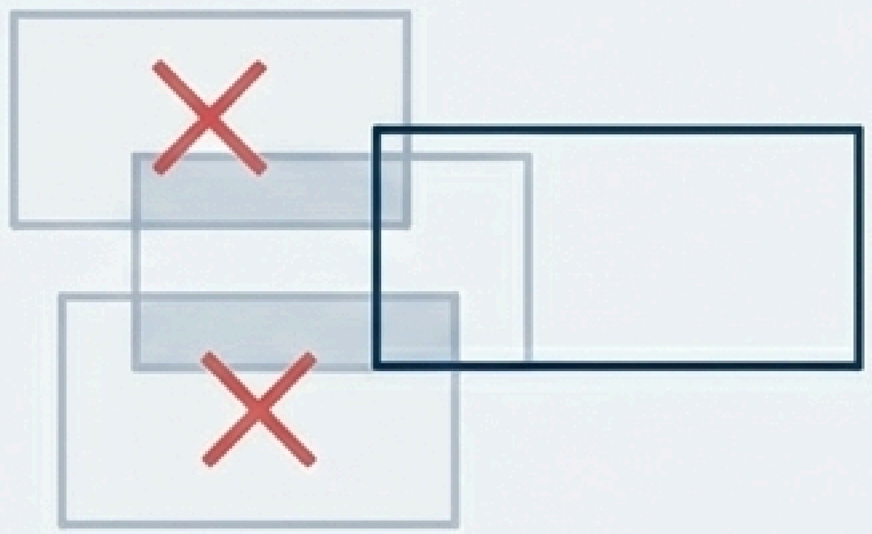
A secondary pass is run on the full, resized image to capture large, macro-scale features like potholes that might be missed in individual tiles.

Pillar 3: Our Solution, Part 2 — Non-Maximum Merging (NMM)

After SAHI creates multiple detections for the same object, we must consolidate them. Standard Non-Maximum Suppression (NMS) is destructive for this task.

Standard NMS (The Problem)

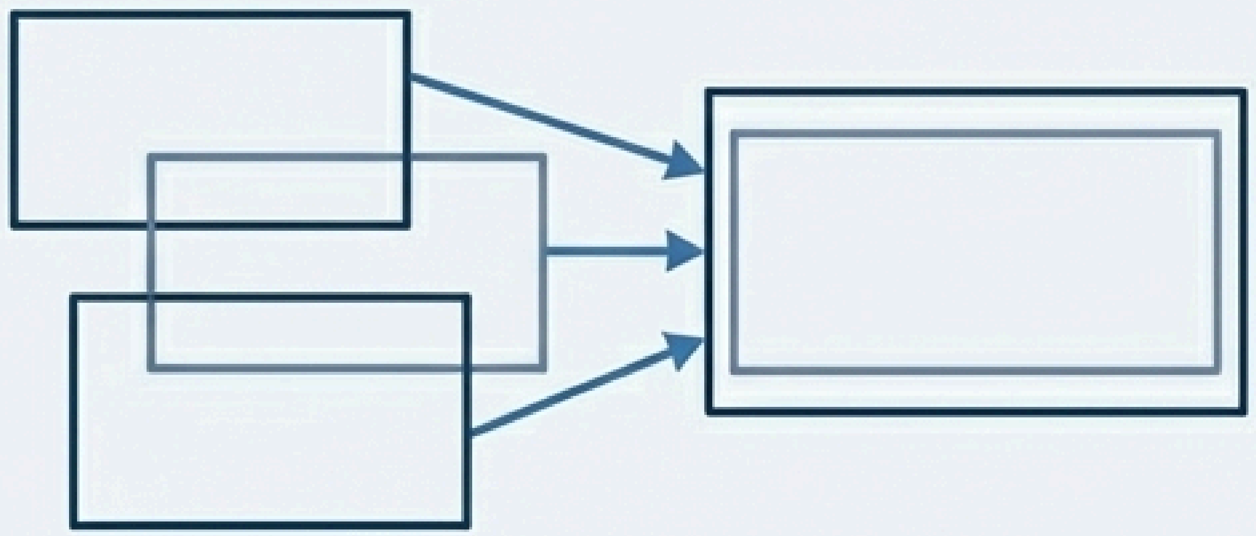
Non-Maximum Suppression **discards** overlapping bounding boxes. For a long crack detected in segments across multiple tiles, this erases parts of the final detection.



Result: Fragmented, incomplete detections.

Our NMM Solution (The Fix)

Non-Maximum Merging **intelligently merges** overlapping boxes. It computes a weighted average of their coordinates to reconstruct a single, continuous bounding box.

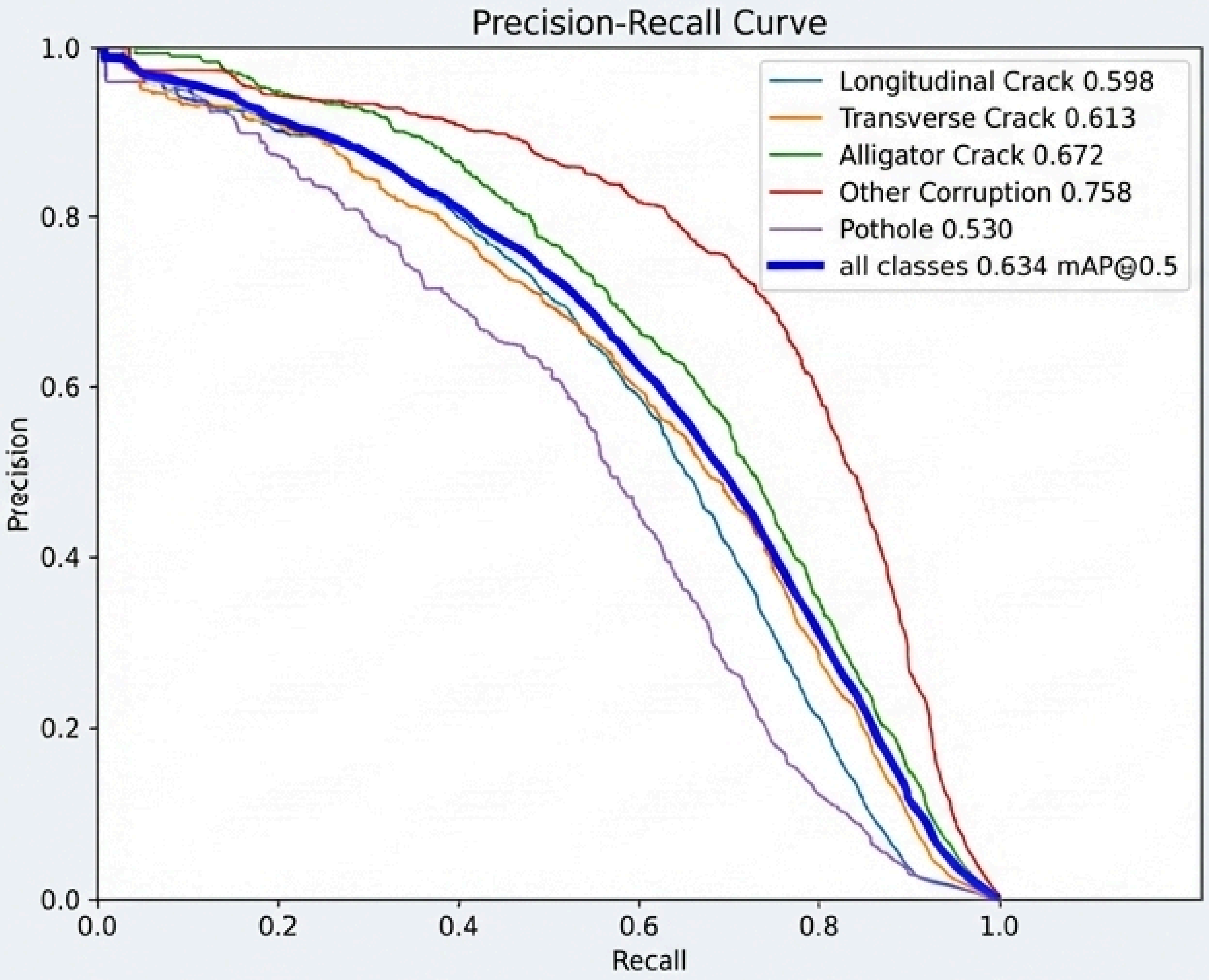


Result: A seamless, unified prediction that accurately represents the full extent of the crack.

NMM effectively stitches together disjointed predictions into a cohesive whole.

Performance Analysis: Quantitative Results

Our final model demonstrates a strong balance between precision and recall, achieving a Mean Average Precision (mAP@0.5) of **0.634**.

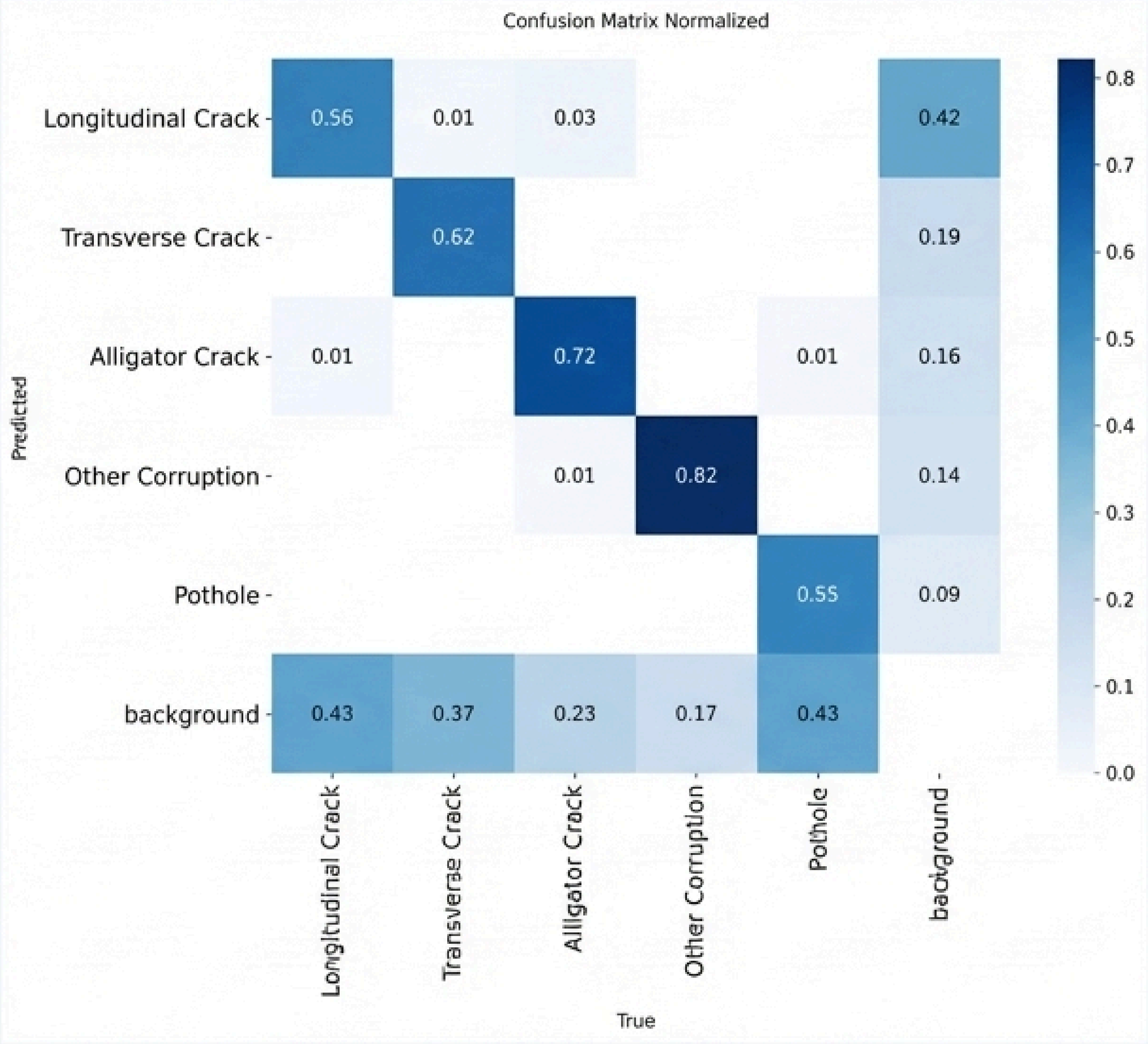


Key Performance Highlights:

- **Overall mAP@0.5:** 0.634
- **Top Performing Classes:** 'Other Corruption' (0.758 mAP) and 'Alligator Crack' (0.672 mAP) show exceptionally high performance.
- **Balanced Performance:** The bold blue 'all classes' line maintains a high-precision curve before dropping, indicating the model is confident and accurate in its primary detections across the board.

Performance Analysis: Deconstructing Model Behavior

The normalized confusion matrix reveals the model's **class-level accuracy** and identifies specific confusion points.



Insights:

- **High True Positive Rates:** The strong diagonal shows high accuracy, especially for 'Other Corruption' (82%), 'Alligator Crack' (72%), and 'Transverse Crack' (62%).
- **Primary Confusion Point:** The model's main challenge is misclassifying "Longitudinal Cracks" as "background" (43% of the time). This is expected, as faint cracks can be visually indistinguishable from the road surface.
- **Minimal Inter-Class Confusion:** Crucially, there is very little confusion *between* different crack types. This validates our choice of the YOLOv8x model for its high discriminative power.

Qualitative Analysis: Resisting False Positives

General vision models often 'hallucinate' damage in ambiguous shadows or textures. Our domain-specific model demonstrates superior precision by avoiding these false positives.

Input Image (000056.jpg)

This image contains challenging lighting and shadows but no actual damage.

Our Model's Prediction

Our model correctly identifies the road as clear, generating **zero false positives**.

Note: The technical report highlights this image as a case where general models fail. Our model's ability to correctly produce no detections is a sign of its robustness and domain-specific training.

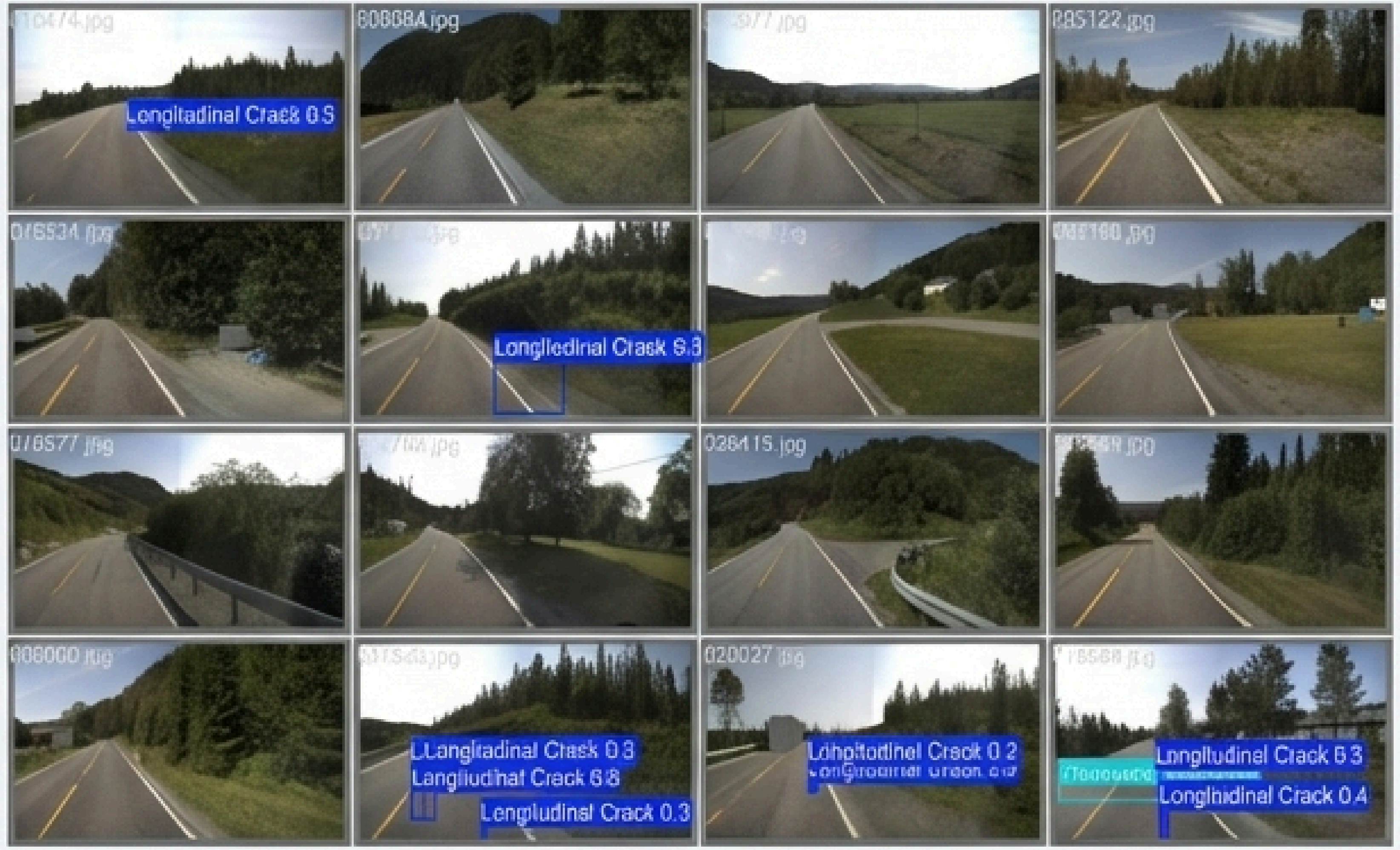
Results in Action: Validation Batch Predictions

This visual comparison on a validation batch demonstrates the model's effectiveness in identifying various damage types in unseen data.

Ground Truth Labels

A sample from the validation set with human-annotated labels.

Our Model's Predictions



The corresponding output from our model. Note the high degree of accuracy in both location and classification, with confidence scores shown for each prediction.

Conclusion: A Robust, Engineered Solution

Our success is the result of a systematic engineering approach that integrates three key elements:

1. **A High-Capacity Model (YOLOv8x):** Chosen for its ability to capture fine-grained textural detail.
2. **A Resilient Training Pipeline**: Optimized with multi-stage augmentation for robust and generalizable learning.
3. **A Specialized Inference Engine (SAHI + NMM):** Custom-built to ensure high-recall detection of features at any scale.

This integrated methodology produced a high-performance model with a **validation mAP of 0.65**, demonstrating a robust and effective solution to the Crackathon challenge.

Thank You.

CYPHERFORCE