Crackathon: Advanced Road Damage Detection via Slicing Aided Hyper Inference and Non-Maximum Merging

Team Name: CypherForce
Members: Vaishnavi Pawar, Vishwajeet Pawar

Table of Contents

1. Abstract

This technical report delineates the engineering methodology for the **Crackathon** challenge. Our proposed solution addresses the high-recall requirements of minute pavement defects while ensuring precision on distinct features such as potholes. The system utilizes a **YOLOv8-XLarge** architecture (~68M parameters), optimized through a disruption-tolerant training pipeline. Deployment is achieved via a novel **Slicing Aided Hyper Inference (SAHI)** mechanism integrated with **Non-Maximum Merging (NMM)**, designed specifically to reconstruct disjointed crack features. The model achieves a mean Average Precision (mAP@0.5) of **0.636** on the validation set.

---

2. Model Architecture Selection

We adopted **YOLOv8x (Extra Large)** following an extensive baseline study comparing it with YOLOv8n (Nano).

2.1 Rationale

Road Damage Detection (RDD) requires the model to distinguish between semantically and visually similar classes, specifically *Longitudinal* versus *Transverse* cracks. Empirical analysis demonstrated that lightweight models lacked the sufficient parameter space to encode these subtle textural variances.

2.2 Architecture Specifications

- **Backbone**: Modified CSPDarknet53 with C2f modules.
- **Parameters**: Approximately 68 Million.
- **Input Resolution**: 640x640 (Training), Tiled 640x640 (Inference).
- **Initialization**: Transfer learning initiated via COCO pre-trained weights to accelerate convergence.

---

3. Training Methodology

3.1 Data Augmentation Strategy

A multi-stage augmentation pipeline was engineered to mitigate overfitting:

- **Mosaic Augmentation (prob=1.0)**: Utilized during the initial 40 epochs to enforce context learning by stitching four inputs into a single frame.
- **MixUp (prob=0.1)**: Blended images to soften decision boundaries.
- **HSV Augmentation**: Hue, Saturation, and Value adjustments (`hsv_s=0.7` , `hsv_v=0.4`) simulated diverse photometric conditions.
- **Close Mosaic Optimization**: Mosaic augmentation was explicitly disabled for the **final 10 epochs**. This enabled the model to fine-tune on natural image statistics, significantly reducing false positive rates for "Other Corruption" classes.

3.2 Convergence Analysis

Analysis of the loss curves revealed a stagnation in mAP@0.5 during the initial three epochs.

- **Attribution**: This volatility was attributed to the aggressive initialization of the Mosaic augmentation.
- **Resolution**: As the learning rate stabilized and the model adapted to the synthetic Mosaic context, performance exhibited a robust, monotonic increase towards the final mAP of **0.636**.

3.3 Technical Standardization

To mitigate environment inconsistencies, we developed and deployed a custom utility, `create_data_yaml.py` . This ensured a standardized data interface across all training nodes, maintaining reproducibility.

---

4. Inference Methodology: SAHI + NMM

Standard scaling methods proved inadequate for detecting sub-pixel cracks in high-resolution imagery. We engineered a custom inference pipeline (`advanced_kaggle_inference.py`) to resolve this.

4.1 Slicing Aided Hyper Inference (SAHI)

The inference mechanism slices high-resolution test images into **640x640** tiles.

- **Overlap Optimization**: An **overlap ratio of 0.40 (40%)** was selected.
    - *Analysis*: A standard 20% overlap frequently bisected linear crack features, resulting in fragmented, low-confidence detections. The 40% overlap ensures complete feature encapsulation within at least one tile.
- **Dual-Pass Strategy**: A secondary inference pass on the full-frame resized image captures macro-scale features (e.g., large potholes).

4.2 Non-Maximum Merging (NMM)

Standard Non-Maximum Suppression (NMS) suppresses overlapping bounding boxes, which is detrimental for continuous features like cracks that may be detected as disjointed segments.

- **Methodology**: NMS was replaced with **Non-Maximum Merging (NMM)**.
- **Mechanism**: NMM computes a weighted average of coordinates for overlapping predictions rather than discarding them. This effectively stitches disjointed segments into a singular, continuous prediction.
- **Parameters**: Match Threshold: `0.5` (IoU), Confidence Floor: `0.15` .

4.3 Robustness

- **Empty File Generation**: The pipeline systematically handles negative samples (undamaged roads) by generating compliant empty files.
- **Normalization**: Coordinates are strictly clamped to the [0, 1] interval.

---

5. Comparison & Qualitative Analysis

Validation against general-purpose Vision models (e.g., Gemini Pro) on challenging samples (Image `000056.jpg`) demonstrated the efficacy of domain-specific training. While general models hallucinated damage in shadow regions, our YOLOv8x model exhibited precise localization with minimal false discover rate.

---

6. Appendix: Configuration Summary

Table 1: Training Hyperparameters

Parameter	Value	Description
Model	`yolov8x.pt`	Extra Large model (68M params)
Epochs	50	Total training duration
Batch Size	24	Optimized for Dual T4 GPUs
Optimizer	SGD	Momentum: 0.937, Weight Decay: 0.0005
LR Scheduler	Cosine Annealing	`cos_lr=True`
Input Size	640x640	Standard square input
Mosaic	1.0 (Epochs 0-40)	Aggressive context augmentation
Close Mosaic	10	Disabled for final 10 epochs
MixUp	0.1	Softens class boundaries
HSV-H	0.015	Hue variation
HSV-S	0.7	Saturation variation
HSV-V	0.4	Value (Brightness) variation

Table 2: Inference Configuration (SAHI + NMM)

Parameter	Value	Rationale
Slice Height/Width	640	Matches training resolution
Overlap Ratio	0.40 (40%)	Prevents splitting of cracks at tile edges
Post-Process Type	NMM	Non-Maximum Merging (vs. NMS)
Match Threshold	0.5	IoU threshold for merging boxes
Confidence Thresh	0.15	Low floor to maximize recall (filtered by NMM)
Standard Pred	True	Full-frame pass enabled for large context

7. Conclusion

This submission represents a systematic engineering approach to Road Damage Detection. By coupling a high-capacity model (YOLOv8x) with a specialized inference engine (SAHI + NMM) and a resilient training infrastructure, we have developed a robust, high-performance solution suitable for the Crackathon challenge.