Bangladesh Road Object Detection for Autonomous Vehicles

DL Enigma 1.0 (SUST CSE Carnival)

Niloy Sarker

Competition Link: https://www.kaggle.com/competitions/dl-enigma-10-sust-cse-carnival-2024

Problem & Dataset

- Autonomous navigation in Bangladesh requires robust object detection across diverse roads and lighting.
- **Dataset**: BadODD – Bangladeshi Autonomous Driving Object Detection Dataset
  - 9 districts, $\sim$9.8K images, $\sim$79K objects, 13 classes (cars, rickshaws, buses, pedestrians, trains, wheelchairs, etc.)
  - Collected with smartphones $\rightarrow$ realistic driving conditions (day/night, motion, occlusion)
  - **Challenges**: class imbalance, varied environments, small/rare objects

Handling Class Imbalance & Lighting Variation

- Dataset had cars & buses dominant, while wheelchairs and other rare classes appeared very few times.
- Applied augmentations:
  - Flip and mosaic to increase variety.
- Used ensemble with **Weighted Box Fusion (WBF)**:
  - Boosted detection of rare classes.

Why Weighted Box Fusion (WBF)?

- Normal Non-Maximum Suppression (NMS) discards overlapping predictions.
- WBF fuses predictions $\rightarrow$ more accurate bounding boxes.
- Helps especially on rare/ambiguous objects.
- Boosted leaderboard score and reduced false positives.
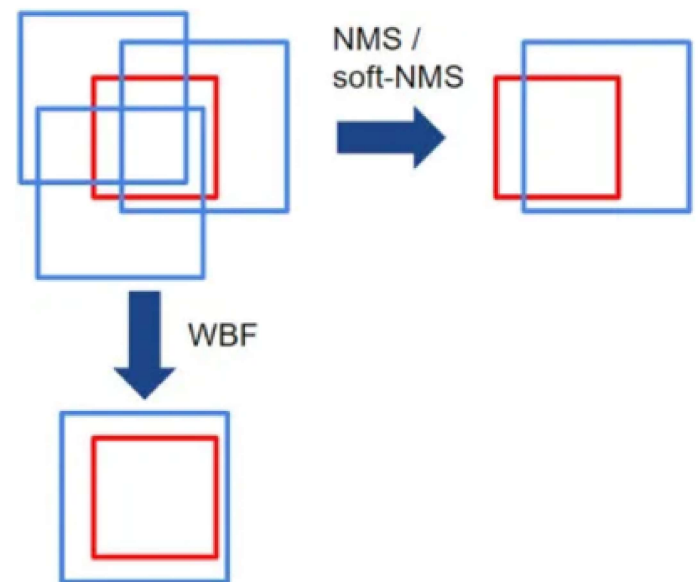
Model Selection

- Traditional detectors like YOLO use convolutional networks.
- DETR introduced transformers for object detection, great at capturing relationships between objects.
- Initially used YOLO, but moved to **RT-DETR** (Real-Time Detection Transformer):
    - Balances accuracy & real-time speed.
    - Handles multi-scale object sizes better.
    - More robust for diverse road conditions in Bangladesh.

Training and Inference Pipeline

- **Framework**: Ultralytics RT-DETR
- **Two models trained**:
  - Model 1 $\rightarrow$ 100 epochs
  - Model 2 $\rightarrow$ 60 epochs
- **Settings**: batch 16, image size 640, LR 0.018
- **Inference Pipeline**:
  1. Two trained models loaded: Model 1 (last.pt, 100 epochs), Model 2 (best.pt, 60 epochs)
  2. Prediction steps:
     1. Run inference on test set.
     2. Collect boxes, scores, classes.
     3. Apply Weighted Boxes Fusion (WBF) $\rightarrow$ merge predictions.
  3. Parameters: weights $= [1,1]$, iou_thr $= 0.4$, skip_box_thr $= 0.001$
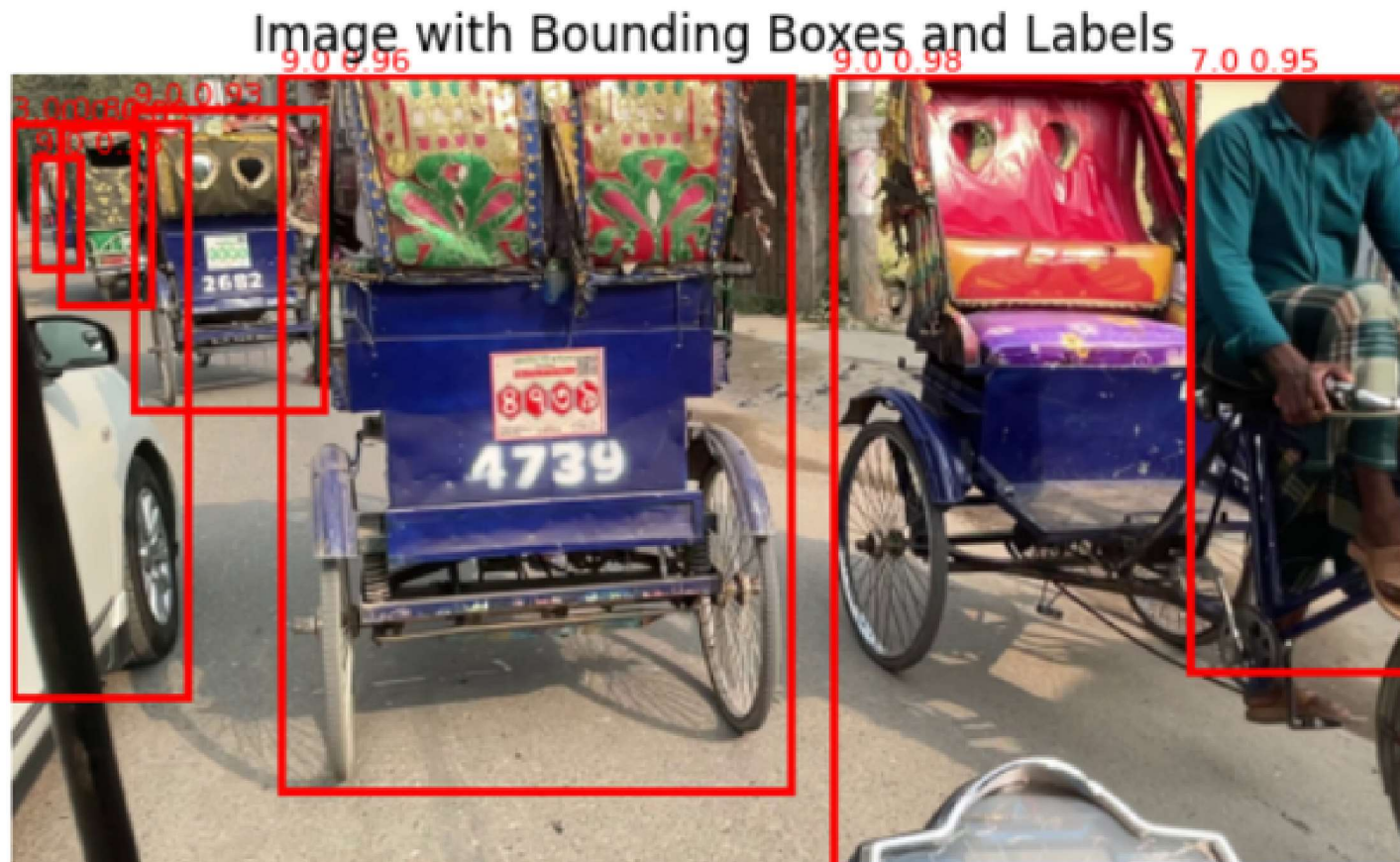
Output Images with Bounding Boxes and Labels

Output Images with Bounding Boxes and Labels

Achievement

- Our team – **SUST Rising Team** – became 1st in SUST due to:
  - Careful model selection (RT-DETR)
  - Weighted Box Fusion (WBF) ensembling
  - Handling rare classes and lighting variation effectively