VISTA26 OBJECT DETECTION – COUNT-AWARE MULTI-PASS INFERENCE PIPELINE

TEAM INFORMATION

Team Name: kalyaniugale

Members:

- **Kalyani Ugale**

- **Darshan Gavate**

Leaderboard score :- 0.97510

EXECUTIVE SUMMARY

Vista26 is a **count-sensitive object detection challenge** where even a **single count mismatch results in a zero score** for that image.

Our solution is specifically engineered for this **strict evaluation behavior**.

Instead of optimizing only **detection confidence**, we designed a **count-aware inference pipeline** that stabilizes object predictions under **dense, overlapping, and occluded conditions**.

Final Results:

- **Vista-like Mean Score:** 0.99700

- **Exact Count Match Rate:** 99.70%

PROBLEM ANALYSIS

Vista26 fundamentally differs from traditional object detection benchmarks.

If the **predicted object count** does not exactly match the **ground-truth count**, the score for that image becomes **zero**.

This transforms the challenge from pure detection into **precise count-aware detection under dense scenes**.

Common Failure Modes

- **Duplicate detections**

- **Missed small or overlapping objects**

- **Over-counting** from recall-heavy inference

- **Under-counting** from strict confidence thresholds

Core Difficulty

The central difficulty is **not detecting objects** ,
it is **predicting the exact number of objects**.

CORE INNOVATION

Detection is easy. Stable counting is hard.

Our primary innovation is a **structured multi-pass inference pipeline** designed specifically to **stabilize object counts**.

The strength of this solution lies in **inference design**, rather than simply increasing model size.

MODEL ARCHITECTURE

- **Backbone:** YOLOv8-Large

- **Training Resolution:** 960

- **Dataset Format:** Converted to YOLO format

High-resolution training improves detection of **small and overlapping objects**.

However, raw predictions remain unstable in dense scenes , requiring **advanced inference reasoning**.

MULTI-PASS INFERENCE STRATEGY

Each image is processed multiple times using different **confidence** and **NMS configurations**.

1 Precision Pass

- High confidence threshold

- Strict NMS

- Removes false positives

2 Balanced Pass

- Moderate confidence threshold

- Stable detection baseline

3 **Recall Pass**

- Low confidence threshold

- Recovers missed objects

4 **Extra Recall Pass**

- Very low confidence threshold

- Used only for difficult images

Each pass produces different object counts.
These variations are intentionally used for **count stabilization**.

DEDUPLICATION BEFORE COUNTING

Bounding boxes are **deduplicated using IoU-based filtering** before counting.

This removes overlapping duplicate detections and significantly improves **count accuracy**.

This step alone provided a major performance gain.

COUNT STABILIZATION USING MEDIAN

We compute the **median object count** across inference passes.

The **median** is robust to:

- Recall overestimation

- Precision under-counting

To prevent recall dominance, the median count is **clipped between Precision and Balanced counts**.

This produces a **stable and realistic target count**.

FINAL SELECTION WITH VOTE BONUS

Candidate boxes are pooled from **Recall** and **Balanced** passes.

Each candidate box receives a **vote bonus** if it overlaps with a **Precision detection**.

Final Score per Box

confidence + vote_bonus

We then select exactly **Top-K detections**, where **K equals the stabilized target count**.

This guarantees:

- ✅ No under-count

- ✅ No over-count

- ✅ High category correctness

PERFORMANCE RESULTS

Evaluated on **1000 labeled test images** using Vista-like scoring logic.

- **Vista-like Mean Score:** 0.99700

- **Exact Count Match Rate:** 99.70%

- **Mean Similarity (when count matches):** 1.00000

- **Runtime:** 199.0 seconds

- **Speed:** 5.03 images per second

Since images with count mismatch automatically score **zero**, these results are highly strict and meaningful.

WHY THIS APPROACH WINS

- Designed specifically for **count-sensitive evaluation**

- Focuses on **count correctness**, not just detection confidence

- Uses structured **multi-pass reasoning** within a single model

- Introduces **deduplication before counting** for dense scenes

- Guarantees exact **Top-K selection**

- Robust, interpretable, and reproducible system

FINAL STATEMENT

This project demonstrates that **deeply understanding the evaluation metric** can outperform simply scaling model size.

Count-aware reasoning is the true key to solving Vista26.