📑 Technical Report: Phase 1 - Model Validation & Architecture Selection

Date: December 2025
Project: CattleCounter (Aerial Livestock Monitoring)
Author: Carlos Luis Noriega - AI Research Engineer

1. Executive Summary

The objective of Phase 1 was to evaluate the feasibility of using Deep Learning to count cattle from aerial drone footage (zenithal view) without prior specific training (Zero-Shot). We compared architectural approaches and established a baseline performance using a pre-trained **Detection Transformer (DETR)**.
Key Result: The baseline model achieved an **84% counting accuracy** (37/44 subjects) in a controlled test environment, demonstrating high viability for production deployment after fine-tuning.

2. Architectural Decision: CNN vs. Transformers

We evaluated two state-of-the-art approaches for the object detection engine:

Option A: CNNs (e.g., YOLOv8)

- **Pros:** Extremely fast inference (Real-time), lightweight.
- **Cons:** Struggles with occlusion and high-density crowds. The "receptive field" is local, making it harder to distinguish individual animals in clumps.
- **Verdict:** Discarded for the batch-processing pipeline, reserved for future edge-only implementations.

Option B: Transformers (DETR - Detection Transformer)

- **Pros:** Utilizes **Self-Attention mechanisms**. This allows the model to understand the

global context of the image. It effectively models relationships between object parts, making it superior for handling occlusion (e.g., separating two cows walking side-by-side).
- **Cons:** Higher computational cost (slower inference).
- **Verdict: SELECTED.** Since the requirements allow for batch processing (asynchronous reporting) rather than millisecond real-time feedback, accuracy is prioritized over raw speed.

3. The "Zenithal View" Challenge

We standardized on a **90° Top-Down (Zenithal) view** for data acquisition.

- **Engineering Advantage:** Eliminates occlusion almost entirely. Each animal occupies a distinct bounding box area on the ground.
- **AI Challenge (The "COCO Bias"):** Standard datasets (COCO) contain images of cows from the side. From above, a cow appears as a rectangular, textured blob.
- **Observations:** The pre-trained DETR model initially misclassified cows as bird, sheep, or bear due to geometric similarities from the top-down perspective.
- **Solution:** We implemented a post-processing heuristic pipeline:
  1. **Class Mapping:** Remapped specific detected classes (bird, sheep, bear) to "Cow".
  2. **Area Filtering:** Discarded detections below $4000px^2$ (noise) and above $150000px^2$.

4. Experimental Results

Test Bench:

- **Input:** 1080p Video @ 60 FPS (Drone flight).
- **Ground Truth:** 44 Cows (Manual Count).
- **Model Prediction:** 37 Cows.
- **Metrics:**
  - **Accuracy:** ~84%
  - **FPS (Inference):** ~15 FPS on CPU (Mac M1/MPS).

Error Analysis:

The missing 16% (7 cows) were attributed to:

1. **Edge Cases:** Cows entering/leaving the frame were filtered out due to low confidence.
2. **Visual Similarity:** Dark cows on dark soil were missed by the pre-trained weights.
3. **Grouping:** The ByteTrack algorithm occasionally swapped IDs when cows moved rapidly in opposite directions.

5. Next Steps (Phase 2)

To bridge the gap from 84% to >98% accuracy:

1. **Data Collection:** Acquire a diverse dataset of 200+ zenithal images.
2. **Fine-Tuning:** Retrain DETR specifically on the new dataset to fix the "Bird/Bear" classification error and improve confidence on dark targets.
3. **Deployment:** Containerize the pipeline for Cloud execution (Azure).