



When I first evaluated the YOLO bounding boxes on a sample video, I noticed that boxes covering chick clusters had a much larger area than boxes covering single chicks. I took all YOLO-predicted bounding boxes in that video, labeled them as either single or cluster, and swept through possible area thresholds from the smallest to the largest box. For each candidate threshold, anything above the threshold was predicted as a cluster box, and anything below the threshold was predicted as a single-chick box. I then chose the area threshold that minimized misclassifications for that video.

I performed a similar analysis to detect overcount / duplicate boxes. In that case, boxes with area below a certain threshold were treated as duplicates/overcounts, and boxes above that threshold were treated as legitimate single-chick boxes. Based on these two thresholds, I defined a simple weighted counting rule for the video:

- Cluster box ( $\text{area} \geq \text{cluster threshold}$ ): contributes +2 to the total count
- Single box ( $\text{duplicate threshold} \leq \text{area} < \text{cluster threshold}$ ): contributes +1
- Duplicate/overcount box ( $\text{area} < \text{duplicate threshold}$ ): contributes +0

Running the video through my YOLO model with this area-based weighting gave a predicted chick count that was within ~2% of the true count (~98% accuracy for that video). A teammate separately analyzed the variance of these predictions across the full video.

The IoU-based refinement came from looking at where the area thresholds were still failing. After labeling boxes, about 13% of them were true cluster boxes, 85% were true single-chick boxes, and around 2% were duplicates/overcounts. With the chosen thresholds, roughly 45% of true cluster boxes were still being predicted as single boxes, because their areas did not exceed the cluster threshold. This makes sense statistically: clusters are a minority class, so they have less influence on the optimal area threshold.

However, when I inspected those misclassified cluster boxes, I saw a pattern: most of them had strong overlap with other boxes—their IoU with a neighboring box was greater than 0.5. Intuitively, this means the model was often covering the same physical “blob” of multiple chicks with several overlapping boxes, each individually too small to be classified as a cluster, but together clearly representing a cluster.

That observation motivated the IoU rule: if a box overlaps more than 50% ( $\text{IoU} > 0.5$ ) with another box, I treat that group as a cluster group rather than a set of independent singles. In practice, this means that for overlapping groups that meet the IoU criterion, I apply the cluster logic and add +2 to the total count instead of +1, because that configuration is much more likely to represent multiple chicks packed together than a single bird.

By combining the area-based thresholds with this IoU-based grouping, the pipeline corrected many of the undercounts from misclassified clusters. After re-running the YOLO model with this updated counting pipeline, the predicted chick count matched the true count exactly (100% accuracy) for that test video. The remaining work now is to quantify how stable that accuracy is across the entire video and across different videos (i.e., analyzing the variance of this method, which we’re currently working on).