Multi-camera detection and tracking for individual broiler monitoring

Thorsten Cardoen[a], Patricia Soster de Carvalho[b], Gunther Antonissen[b], Frank A. M. Tuyttens[c,d], Sam Leroux[a], Pieter Simoens[a]

[a]*IDLab, Department of Information and Technology, Ghent University - imec, Technologiepark-Zwijnaarde 126, Ghent, 9052, Belgium*
[b]*Department of Pathobiology, Pharmacology and Zoological Medicine, Ghent University, Salisburylaan 133, Merelbeke, 9820, Belgium*
[c]*Flanders Research Institute for Agriculture, Fisheries, and Food (ILVO), Melle, 9090, Belgium*
[d]*Department of Veterinary and Biosciences, Ghent University, Salisburylaan 133, Merelbeke, 9820, Belgium*

Abstract

Welfare concerns in poultry farming have driven the need for advanced monitoring solutions to study broiler activity and health. However, existing research predominantly relies on single-camera setups, which are prone to occlusions from equipment such as feeders and lighting, limiting their effectiveness. To address this, we propose a multi-camera setup that enables comprehensive broiler localization and tracking from a top-down view of the pen. To support this approach, we introduce MVBroTrack[1], an open-source dataset containing real-world data with annotations for various subtasks critical to broiler studies. We demonstrate robust performance of our multi-view detection pipeline throughout the six-week broiler lifespan despite significant changes in visual appearance. Additionally, we present a novel unsupervised tracking method that surpasses the traditional tracking by detection paradigm, improving the IDF1 score by 3% and increasing the proportion of mostly tracked broilers by 5%. By reducing the need for manual observation, our multi-camera pipeline facilitates exhaustive studies of broiler behavior and welfare, paving the way for significant advancements in poultry research and farming practices.

[1]`https://github.com/decide-ugent/multi-view-broiler-tracking.git`

Keywords: Multi-camera detection, Object tracking, Broiler welfare monitoring, Sensor fusion

1. Introduction

Effective monitoring of broiler chickens is essential for understanding their behavior and ensuring welfare standards (Aydin, 2017). To this end, various sensor-based technologies have enabled automated methods for monitoring broilers in pens (Ojo et al., 2022). Among these, cameras offer high-fidelity data for a reasonable infrastructure cost. Video-based observation offers a non-invasive method to study broiler behavior, allowing researchers to analyze natural patterns without disturbing the animals. However, in practice, many studies still rely on manual video analysis (Aldridge et al., 2022; Baxter et al., 2018, 2019; Trocino et al., 2020; du Plessis et al., 2021), which is time-consuming and prone to observer variability. This manual approach often limits both the duration and scope of behavioral studies, making it challenging to detect subtle behavioral changes that could indicate important welfare or behavioral patterns.

These limitations in manual analysis, combined with the increasing demand for comprehensive behavioral monitoring, have driven the development of automated video analysis techniques for broiler welfare assessment. Earlier techniques employed a grid-based/density-based approach to estimate occupation scores for different zones within a pen (Peña Fernández et al., 2018). However, as these methods only provide aggregated information, research has shifted from the group level to the individual level (Li et al., 2020b) where the exact positions of individual broilers are determined (Guo et al., 2020; Novas and Usberti, 2017; Li et al., 2020a). These methods relied on traditional computer vision techniques, involving grayscale conversion, blurring, Otsu thresholding for segmentation, and region enhancement through dilation.

While these methods have shown some efficacy in very controlled settings, they are limited in their adaptability to various environmental conditions. Recent advancements in machine learning, particularly deep learning object detection models like YOLO (You Only Look Once) (Jocher et al., 2022; Zaidi et al., 2022; Neethirajan, 2022), have paved the way for more robust and flexible broiler detection systems (Li et al., 2022; Van der Eijk et al., 2022; Yang et al., 2022; Cao et al., 2021), offering improved accuracy and

versatility compared to traditional techniques that struggle with complex backgrounds and lighting variations.

These monitoring systems utilize single-camera setups which have a limited field of view, making it challenging to monitor the entire pen. Moreover, due to the frequent occlusions by other animals, drinking or feeding equipment (Guo et al., 2021; Okinda et al., 2020) the detection accuracy of single camera setups is degraded and can lead to incomplete results. Using multiple cameras can eliminate blind spots and reduce occlusions while the incremental cost of additional cameras to an existing set-up is mostly limited to the cost of the camera. This enhances the ability to monitor the entire flock effectively.

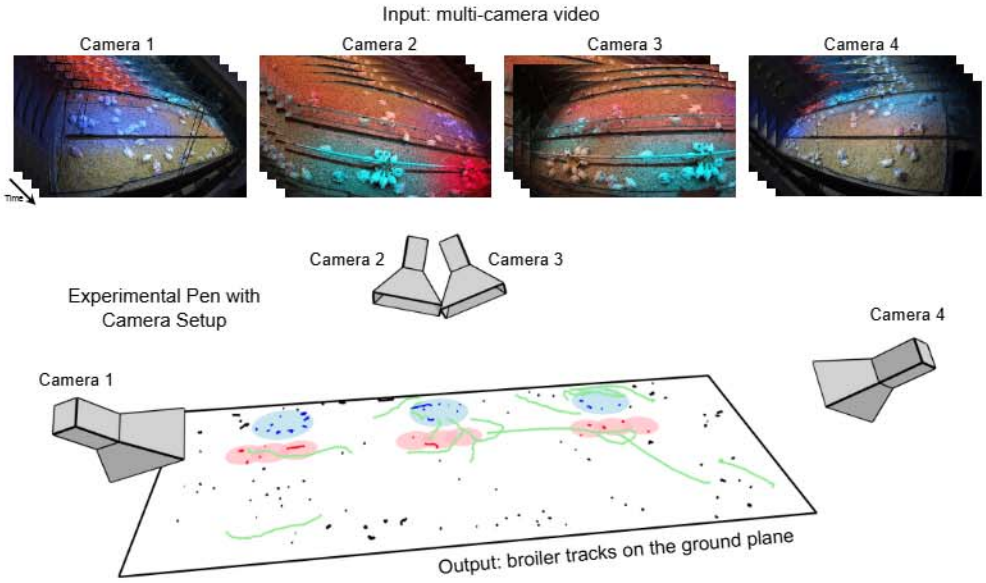

Figure 1: Overview of the multi-camera tracking system for broiler behavior analysis. The 3D visualization shows the experimental pen setup with four overhead cameras, feeders (light blue circles), and drinkers (pink circles). Colored lines indicate bird tracks: green for active movement, black for inactivity, blue for feeding, and red for drinking. Utilizing multiple cameras enables comprehensive monitoring and analysis of poultry behavior, movement patterns, and resource utilization within the pen environment. Figure is best viewed in color.

Multi-camera setups are standard in pedestrian tracking systems but remain largely unexplored for broiler tracking. Recent multi-view pedestrian detection and tracking methods (Hou et al., 2020; Hou and Zheng, 2021;

Hwang et al., 2022; Vora et al., 2023), leveraging pedestrian datasets (Ferryman and Shahrokni, 2009; Chavdarova et al., 2018; Hou et al., 2020; Han et al., 2023), project visual features directly to the ground plane, producing detection maps without requiring initial object detection algorithms. While being effective, it depends on datasets with ground plane pedestrian positions which require labelling across multiple cameras for which there is a lack of tools. Alternatively, projecting 2D detections to the ground plane (López-Cifuentes et al., 2022; Lima et al., 2022) only requires bounding box annotations, which are easier to obtain using readily available tools like Tzutalin (2015); Cartucho et al. (2018); CVAT.ai (2024).

While accurate detection can provide valuable insights into broiler preferences, tracking is indispensable for effective monitoring. Tracking allows continuous observation of individual broiler over time, providing information into their activity levels (Campbell et al., 2024). Tracking methods typically follow a tracking-by-detection paradigm (Engilberge et al., 2023; Chavdarova et al., 2018; Campbell et al., 2024), associating multi-view detections across time (Wang et al., 2019; McLaughlin et al., 2015), or use tracking labels to learn features (Teepe et al., 2024). However, adapting appearance-based tracking is challenging for animals (Zhang et al., 2023), as they lack distinguishing features like pedestrian clothing.

In this paper, we present a multi-view detection and tracking system designed for precision and adaptability. Using four synchronized and meticulously calibrated cameras, we capture a comprehensive overview of the pen, see Figure 1. The pipeline begins with an object detection model trained on our custom dataset to generate per-camera bounding box detections. Note that this is the only part in our framework that requires supervision. Currently, the state-of-the-art in video-based broiler detection and tracking is limited to chickens of at least two weeks old (Guo et al., 2020; Yang et al., 2022; Van der Eijk et al., 2022; Novas and Usberti, 2017; Li et al., 2022). Our Multi-View BROiler Tracking (MVBroTrack) dataset extends this and encompasses detection and tracking throughout the broiler life cycle, from 5 days to 38 days old.

For multi-view detection, these bounding box detections are then fused across views to create a ground-plane detection map, where we introduce a heuristic to approximate the broilers' feet locations. This multi-view detection process, which achieves 85.8% Multi-Object Detection Accuracy (MODA) and 93% recall, is evaluated independently, as it provides a computationally efficient solution for estimating broiler distributions when detailed tracking

is not required. For more detailed analyses, such as assessing activity levels, tracking is required. Typically, the ground plane detection maps from the previous step are linked together temporally to obtain tracks. However, this tracking by detection paradigm is heavily impacted by inconcistencies of the detection maps across different timesteps. Our tracking method addresses this problem by first associating bounding box detections across time in the image plane which improves the information for the subsequent fusion step. We introduce a novel tracklet fusion algorithm to fuse the tracklets originating from different cameras on the ground plane both spatially and temporally. Compared to the traditional tracking-by-detection paradigm, our method demonstrates notable improvements, increasing the Identity F1 score (IDF1) by 3% and the proportion of mostly tracked broilers by 5%, highlighting the effectiveness of our approach.

The main contributions of this paper are summarized as follows:

- **Introducting multi-view setup for broiler monitoring:** We extend single-camera broiler detection pipelines to fuse multi-view information, addressing occlusion challenges and enabling comprehensive pen coverage.
- **Adaptability across growth stages:** Despite significant changes in broiler appearance as they grow, our system maintains high detection accuracy, achieving an average MODA of 85.8% and 93% recall.
- **Improved tracking:** We introduce a novel tracklet fusion approach, which achieves a 3% higher IDF1 score while increasing the amount of mostly tracked broilers by 5% compared to a traditional tracking-by-detection pipeline.
- **Open-access resources:** We provide our MVBroTrack dataset and code to facilitate further research and improvements in the domain.

2. Materials and methods

2.1. Multi-camera broiler dataset

The following section describes our Multi-View Broiler detection and tracking dataset MVBroTrack. This includes the data collection which details the pen setup and camera configuration, the annotation procedures and the metrics used for evaluating the performance of our pipeline.

2.1.1. Data Collection

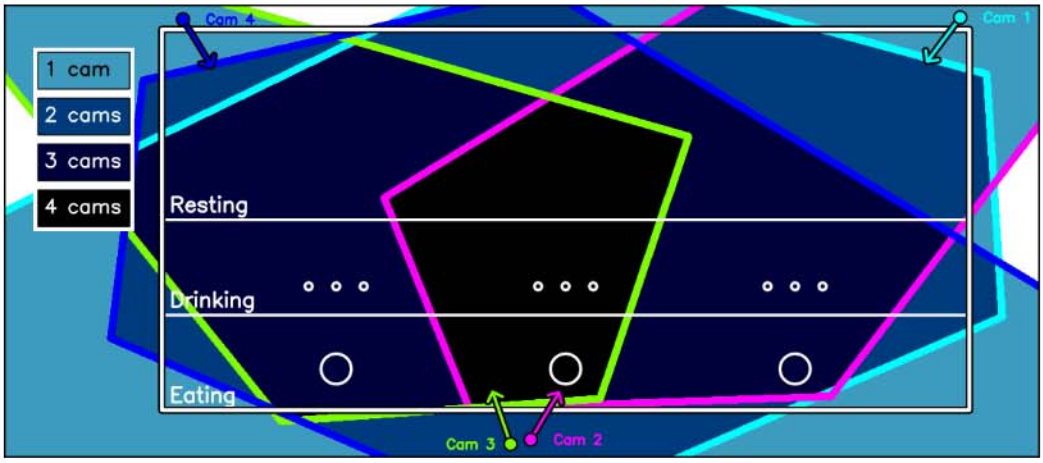

Figure 2: Visualization of camera placement and coverage in the experimental pen. The pen contour is shown in white. The four arrows indicate the direction and location of each camera. Each arrow has a color and corresponds to the contour of the visible area for that camera. The blue shades represent the degree of camera overlap.

This study is part of the WISH research project (ILVO and imec, 2022-2024) which aims to automate welfare assessment using multi-modal sensor data and test the impact of various pen infrastructure conditions, such as lighting color. The experiment consisted of 4 rounds, each involving 560 male Ross 308 chickens, totaling 2240 birds. Each round presented four pens of a size of 36 m^2 (9x4 m). A total of 140 one day old chicks were randomly selected and placed in each pen. All animals received the same starter (from day 0 to 9), grower (from day 10 to 22) and finisher (from day 23 to 41) food and water ad libitum. The pens were logically (but not physically) divided into zones for eating (three feeders), drinking (three sets of three drinkers), and resting, see Figure 2. Stocking density was 3,8 birds/m^2 on purpose so each bird could freely move around and present normal behaviour patterns. Lighting and temperature conditions were managed in accordance with the breed's management manual.

We mounted four cameras on the walls around each pen, selecting positions from a range of possible mounting locations. The chosen configuration, shown in Figure 2, provided the most effective coverage of the entire pen. Our video recording configuration featured 16 Hikvision DS-2CD2143G2-I cameras paired with a Hikvision DS-7616NI-K2/16P 16-channel NVR con-

nected via Power-over-Ethernet. Data was stored on WD Purple 5 TB hard drives. The cameras captured video at a resolution of 1920x1080 pixels at 25 fps and a bitrate of 8120 bits per second. We recorded in intervals of one or two hours, resulting in an average of 837 hours of video per pen.

2.1.2. Data Annotation

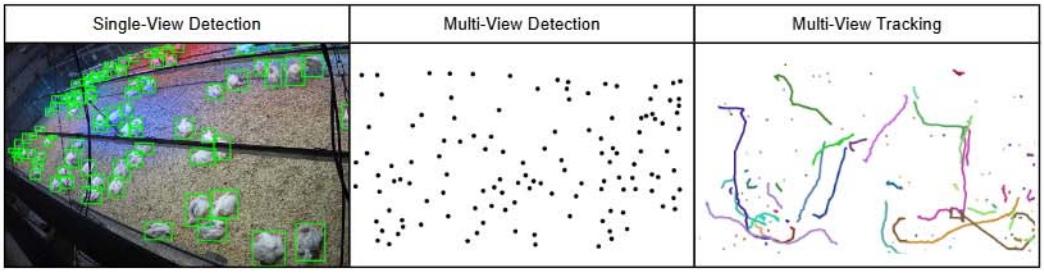

Figure 3: This figure shows one dataset example per subtask. Single-view detection encompasses bounding box labels within the image plane while multi-view detection and tracking are points and tracks on the ground plane respectively. Note that single-view detection labels are used for training while multi-view detection and tracking are only used for hyperparameter tuning and evaluation.

In this section, we provide an overview of annotations used for evaluating each subtask of the broiler tracking and detection pipeline. Table 2 presents a comparative summary of labels collected across different growth stages and labelling types. Recognizing the distinct physical transformations between young and mature broilers, we categorized each set into the three dietary phases discribed in the previous section: starter, grower, and finisher.

In Table 1 the characteristics of broilers at different growth stages are shown. The transition from highly active younger broilers to the less active behavior of older birds as well as the increase in size and density over time (broilers grow but pens stay the same size) lead to unqiue challenges across the different growth stages. The decrease in activity over time may be associated with the rapid increase in body weight Bizeray et al. (2000), rising from 0.288 kg on day 9 to 1.131 kg on day 22, and reaching 3.110 kg by day 41 Aviagen (2022). Due to genetic selection, studies have shown that bone development does not keep pace with the rapid muscle growth in fast-growing chickens Santos et al. (2024), which can impair leg health Fernandes et al. (2012). Additionally, litter quality tends to deteriorate over time, increasing the incidence of problems such as footpad lesions and hock burns Peña Fer-

Table 1: Characteristics of broilers at different growth stages.

Phase	**Starter**	**Grower**	**Finisher**
Ages	Days 0-9	Days 10-22	Days 23-41
Activity Level	Very high activity; frequent movement	Moderate activity; more structured movement patterns	Low activity and limited mobility. Longer periods sitting and inactive
Appearance Features	Small size; yellow down feathers; less distinct body shape	Medium size; developing white feathers; more defined body shape	Large size; fully developed white feathers; pronounced body shape
Detection Difficulties	Challenging due to small size and rapid movements, can easily blend in with the background	Better defined features but still active, which can lead to motion blur	Easier detection due to size and low activity level. However, occlusion more frequent due to their bigger size and higher density of the pen
			

nández et al. (2018), which may further contribute to reduced activity levels Riber and Wurtz (2024).

In order to fuse information across views, calibration parameters for each camera are needed. Mutli-camera calibration was performed to ensure accurate projective geometry between views, this calibration procedure is detailed in Appendix A.

Over the recording span of six weeks, we selected 235 images at periodic intervals from different cameras to ensure enough spatial and temporal diversity. These bounding box annotations are used to train and evaluate detection models and where obtained using the OpenLabeling tool Cartucho (2018).

To evaluate the multi-view detection performance, the dataset includes 51 sets of four frames with annotated ground plane detections, including ages from 5 to 38 days. Using custom tools, we indicated the feet position of each broiler in each image plane, these feet positions are subsequently given a unique ID and projected onto the ground plane. Unique IDs are are combined into a point set if they originate from the same broiler. Lastly an arithmic mean is computed to obtain the ground truth ground plane detections.

To evaluate the tracking performance, the dataset includes tracks on the ground plane for 816 individual broilers. Similarly to the multi-view detection dataset, feet locations of each broiler in each camera view are first manually tracked using the CVAT annotation tool CVAT.ai (2024). These individual camera tracks are then projected onto the ground plane. To ensure consistent tracking across multiple views, we manually combined the track IDs of matching broilers across different cameras. After assigning a unique track ID to each broiler on the ground plane, integrating track IDs from different cameras, we obtained the ground truth ground plane tracks by computing the arithmic mean for points originating from different cameras at the same timestep. This process was highly labor-intensive, as approximately 70 broilers are visible per camera view, resulting in around 140 tracks on the ground plane.

In Figure 3 we show an example of each label type of our dataset. On the left we show the bounding box labels in the image plane for the single-view detection task. In the middle we show the multi-view detection task with localisations for each broiler on the ground plane. On the right we show tracks for each broiler with a unique ID over time on the ground plane.

Table 2: An overview of the MVBroTrack dataset: Labels across different growth stages and labelling types

Task	Stage	Frames	Broilers	Image Plane	Ground Plane	
				BBoxes	Points	Tracks
Single-view detection	Starter	56	4106	4106	-	-
	Grower	65	5714	5714	-	-
	Finisher	114	11414	11414	-	-
	Total	235	21234	21234	-	-
Multi-view detection	Starter	7x4	893	-	893	-
	Grower	21x4	2791	-	2791	-
	Finisher	23x4	3022	-	3022	-
	Total	51x4	6706	-	6706	-
Multi-view tracking	Starter	1423x4	275	-	-	275
	Grower	1494x4	276	-	-	276
	Finisher	1439x4	265	-	-	265
	Total	4356x4	816	-	-	816

2.2. Methods

In this section, we present an overview of the multi-camera multi-object tracking pipeline, as illustrated in Figure 4. The first subsection covers the single-view object detection process, which serves as the foundation for both detection and tracking methods by generating per-camera, per-timestep bounding box detections. The second subsection details the multi-camera detection pipeline that produces a ground-plane detection map per timestep. Lastly the mutli-view tracking pipeline is discussed, detailing our novel tracking approach and the comparison to the traditional tracking-by-detection-matching (TBDM) approach.

2.2.1. Single-View Detection

The goal is to monitor broiler welfare throughout their entire lifespan in the pen. This entails addressing the challenge of detecting chickens of varying ages, each with their own challenges (see Table 1). We adopt YOLOV11 object detection, a robust and efficient method capable of handling object size variations, providing a versatile lifelong detection solution. We fine-tune the pre-trained YOLOV11 model (Jocher et al., 2022) on our dataset to obtain per-view bounding boxes for each broiler. Both of the subsequent

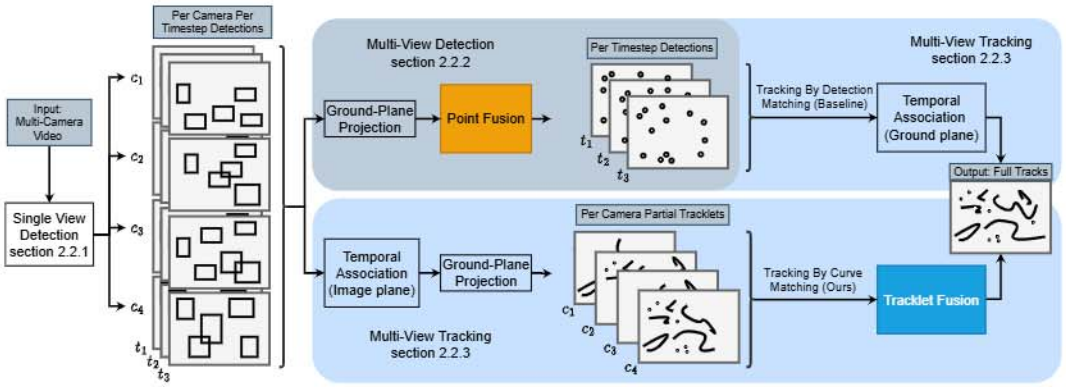

Figure 4: This figure highlights the structural differences between the traditional tracking-by-detection-matching (TBDM) method and our novel tracking-by-curve-matching (TBCM) approach. The top path illustrates the TBDM method, where per-camera detections are projected onto a ground plane, followed by point fusion creating per timestep ground-plane detections, and temporal association to produce ground-plane tracks. The bottom path presents our proposed method TBCM, introduced in this paper. In TBCM, partial temporal association occurs first generating per-camera tracklets, followed by ground-plane projection. The tracklets on the ground plane are then combined through tracklet fusion to yield the final output of ground-plane tracks.

tracking pipelines utilize these per timestep, per camera bounding boxes. Training and testing the YOLO models was done using an NVIDIA A100 80GB GPU with the auto-batching feature enabled. Each model was trained for 200 epochs and the best performing model was selected based on the validation set performance.

It is important to note that this detection component is designed to be modular and interchangeable. While YOLOV11 is employed in this study, the pipeline is compatible with alternative or future object detection models. Integrating newer or more advanced models can further enhance the accuracy and robustness of the entire tracking system without requiring modifications to the overall framework.

2.2.2. Multi-View Detection

The position of each broiler on the ground plane can be approximated by the midpoint between its feet, however due to the mounting of cameras at elevated positions, the chicken feet are often not visually discernible. For this reason a geometric approach is needed to determine which point in the bounding box should be projected onto the ground plane. The problem is

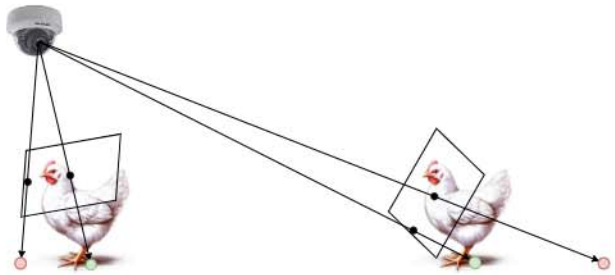

Figure 5: Illustration of the challenge for estimating the point within the object bounding box corresponding to the position between the two feet. As visualized, always selecting the center of the bounding box is not always the best option. Instead, we employ a point selection procedure that is based on the distance to the camera.

illustrated in Figure 5. To find the point within the bounding box that corresponds best to the broiler's position, the following options were explored: the center of the bounding box, the bottom center of the bounding box (which, in general, is closer to the ground for chickens further away from the camera) or a linear interpolation between the bottom center and the center depending on the distance from the camera. According to the analysis in Appendix C, linear interpolation gave the most accurate positioning. This approximation, however, does not take into account whether the broilers are sitting or standing. Because the majority of broilers are sitting down, the linear interpolation is tuned for these and consequently does not work as well for standing broilers. We adres this issue with our tracklet fusion approach described in the following section. This processing step from bounding box detections to per-view selected points is the first step of the multi-view detection pipeline, see Figure 6.

Using the camera's intrinsic and extrinsic parameters, we can project the detections of each camera to the ground plane. We discard points that fall outside the boundaries of the pen. We also discard points beyond a preconfigured distance from the camera to remove the negative influence of noisy bounding boxes at the extremities of the camera's view. In section 3, we discuss its importance to the performance of the pipeline. The purpose of point fusion is to find the projected points from each camera image that correspond to the same animal and to combine them into a final position estimation. For this purpose, we adopt the fusion method described in López-Cifuentes et al. (2022); Zhu (2019), which we refer to as point fusion in Algorithm 1 and Algorithm 2. Note that D in the following algorithms can refer to either

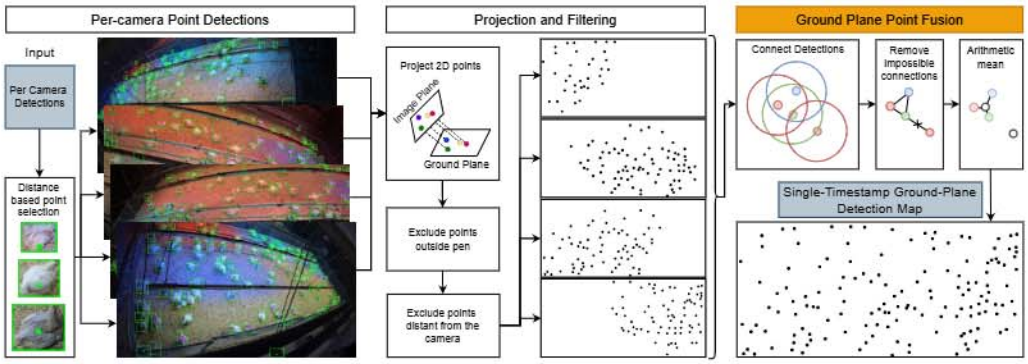

Figure 6: Multi-View Broiler Detection Pipeline: The pipeline commences with performing object detection on time-synchronized input frames from $N = 4$ cameras. For each bounding box, the midpoint between the broiler feet is approximated and projected onto the ground plane. The subsequent geometric fusion of multi-view detections culminates in a unified detection map, effectively synthesizing information from all four input images.

point or tracklets (discussed in the following section). They are shown together only to highlight the differences beween each other. To combine the broiler detections from each viewpoint, connected components are created in a graph representation. Detections are connected if two conditions are met. The first condition is that the projected points have to originate from different cameras and the second condition is that their Euclidean distance has to be lower than the fusion threshold. Lastly the arithmetic mean of the points is calculated to obtain the location of the broilers detected from multiple viewpoints. We set the fusion radius to be twice the distance threshold R, see section 2.2.4, for calculating the metrics so that two detections from different cameras that lie within the radius for determining a True Positive (TP) can still be combined, see Figure 7.

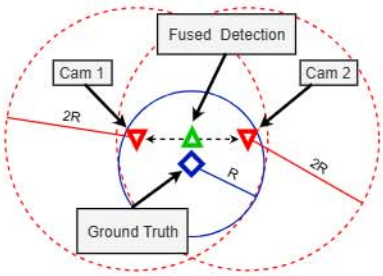

Figure 7: Visualisation of the difference between the fusion radius 2R and the radius R used for determining a true positive. The ground truth is indicated with ◇, the single-camera detections using ▽ and the final prediction (arithmetic mean) using △. Figure is best viewed in color.

Algorithm 1 Graph Construction

procedure BUILDGRAPH(D) ▷ D: detection or tracklet set
  $G \leftarrow$ empty graph
  for all $d_i \in D, d_j \in D$, s.t. $j > i$ **do**
    $r \leftarrow \infty$ ▷ initialize distance
    if $d_i.camera \neq d_j.camera$ **then**
      $r \leftarrow \text{EuclideanDistance}(d_i, d_j)$ | Point Fusion
    end if
    $I \leftarrow d_i.t \cap d_j.t$ ▷ temporal overlap
    $p_1 \leftarrow \text{end\_point}(\text{earlier}(d_i, d_j))$ ▷ end point of earlier track
    $p_2 \leftarrow \text{start\_point}(\text{later}(d_i, d_j))$ ▷ start point of later track
    $K \leftarrow p_2.t - p_1.t$ ▷ temporal distance
    if $d_i.camera \neq d_j.camera$ **then**
      if $I \neq \emptyset$ **then**
        $r \leftarrow \text{FréchetDistance}(d_i|_I, d_j|_I)$
      else if $K \leq \text{max\_gap\_threshold}$ **then**
        $r \leftarrow \text{EuclideanDistance}(p_1, p_2)$ | Tracklet Fusion
      end if
    else
      if $I \neq \emptyset$ **then continue**
      else if $K \leq \text{max\_gap\_threshold}$ **then**
        $r \leftarrow \text{EuclideanDistance}(p_1, p_2)$
      end if
    end if
    if $r \leq R$ **then**
      Add edge (d_i, d_j) to G with weight r
    end if
  end for
  return G
end procedure

Algorithm 2 Multi-Camera Fusion

```
1: procedure FuseDetections(D, G)
2:     found ← ∅
3:     results ← ∅
4:     D_sorted ← Sort(D, key = d_x)                                  Point Fusion
5:     D_sorted ← Sort(D, key = [-temporal_length, d_x0])             Tracklet Fusion
6:     for each detection d in D_sorted do
7:         if d ∉ found then
8:             temp ← [d]
9:             Add d to found
10:            while d ∈ G.nodes do
11:                Q ← ∅                                              ▷ Candidates
12:                for each neighbor d′ of d in G do
13:                    if d′ ∉ found and d′ ∈ Neighbors(p) for any p ∈ temp then
14:                        if d′.camera ≠ p.camera for all p ∈ temp then      | Point
15:                            Add (r_d′d, d′) to Q                          | Fusion
16:                        end if                                            |
17:                        if not (d′.camera = p.camera & d′.t ∩ p.t ≠ ∅)    |
18:                                                for any p ∈ temp then     | Tracklet
19:                            I ← d′.t ∩ d.t                                | Fusion
20:                            Add (r_d′d, |I|, d′) to Q                     |
21:                        end if                                            |
22:                    end if
23:                end for
24:                if Q = ∅ then
25:                    Add temp to results
26:                    break
27:                end if
28:                Q_sorted ← Sort(Q, key = r)                    Point Fusion
29:                Q_sorted ← Sort(Q, key = [−|I|, r])            Tracklet Fusion
30:                d_best ← Q_sorted[0]
31:                Add d_best to temp and found
32:                d ← d_best
33:            end while
34:        end if
35:    end for
36:    return results
37: end procedure
```

2.2.3. Multi-View Tracking

Multi-view tracking involves the association of the single-view detections both temporally and spatially. We investigate both Simple Online and Real-time Tracking (SORT) (Bewley et al., 2016) and multiple Successive Shortest Paths (muSSP) (Wang et al., 2019) for connecting detections temporally. SORT is a real-time approach that considers only past detections, while muSSP is a global method capable of leveraging both past and future detections for temporal association. Both methods can operate on point-based inputs (x, y) or bounding boxes (x, y, width, height) and are unsupervised, requiring no training on labeled trajectories. Given the limitations of re-identification (reID) features in animal tracking due to the high similarity in appearance among individuals (Zhang et al., 2023), we opted not to use advanced methods reliant on these features. Additionally, we avoided trackers requiring training on labeled trajectories as obtaining these labels is very costly.

Given the fused detections on the ground plane from section 2.2.2, a straightforward way to track the broilers is to connect the ground plane detection maps (which are already fused spatially) across multiple timesteps. As detection is a prerequisite for this approach, this is often denoted as the tracking-by-detection-matching (TBDM) principle (Campbell et al., 2024). This approach however is heavily impacted by inconsistencies across multiple timesteps. To combat this, we propose to fuse tracklets both spatially and temporally at the same time. Instead of using the trackers, described above, directly on the ground plane detection maps, they are now instead used within each camera seperately, using bounding boxes as input instead of points on the ground plane. After this step, the bounding box detections are partially fused temporally, which are called tracklets.

To process the tracklets, each tracklet is projected onto the ground plane, using the same procedure described in the previous section on multi-camera detection, taking into account the distance to the camera. Once the tracklets are projected, tracklets too far away from the camera or outside the boundaries of the pen are filtered out. A key challenge in this process is determining the mid-feet position of the broilers, as this position varies significantly depending on whether they are standing or sitting. As discussed in section 2.2.2, the linear interpolation of the bounding box is used to approximate the mid-feet position and works well for broilers that are sitting down. Using tracklets, we can determine the movement of each broiler by

thresholding the Euclidian distance between the first and last point of each tracklet. This allows us to identify and adjust for broilers that are standing up, which would be projected too far away from the camera as the mid-feet position approximation is optimized for broilers that are sitting down. By shifting the tracklets towards the camera depending on the size of the broiler, see Figure 9, we can ensure the projected tracklets are corrected for moving broilers, making the data more reliable for further processing.

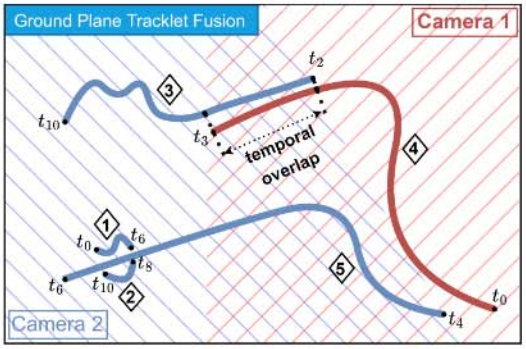

Figure 8: This figure illustrates the four cases when adding two tracklets to the graph. The blue and red hatched regions represent the visible area of the respective cameras. Black points represent specific timesteps while the diamond shapes represent the projected per camera tracklet IDs.

Table 3: Addition to the graph depends on multiple decision criteria. Do the tracklets come from the same camera (SC) and is there temporal overlap (TO) between the tracklets. Depening on these criteria, the algorithm needs to compare (TC) either curves (cur) or points (pts).
A: Euclidean dist $< th_{spatial}$ & temporal dist $< th_{temporal}$
B: Fréchet dist $< th_{spatial}$

IDs	SC	TO	TC	add
1, 2	✓	✗	pts: $1t_6$, $2t_8$	✓ if A
4, 5	✗	✗	pts: $4t_3$, $5t_4$	✓ if A
1, 5	✓	✓	✗	✗
3, 4	✗	✓	cur: 3, 4	✓ if B

To fuse tracklets across space and time, we consider four cases to build up the graph representation, see Figure 8, Table 3 and Algorithm 1. If two tracklets are added to the graph, they represent a potential match. Unlike the fusion of points on the ground plane, where detections cannot originate from the same camera, tracklets can originate from the same camera if there is no temporal overlap. In this case, if both the temporal distance and spatial distance between the last point of the earlier tracklet and the first point of the later tracklet are smaller than the predefined thresholds, these tracklets can be potentially fused and are added to the graph. The same conditions apply for tracklets from different cameras without temporal overlap. In the third case, when tracklets originate from the same camera but have temporal overlap, they cannot be merged because the same broiler cannot appear in

two tracks in the same timestep. Lastly, when there is temporal overlap and the tracklets originate from different cameras, e.g. tracks 3 and 4 in Figure 8, we measure the distance between the two tracklets over the overlapping temporal segment using the Fréchet distance. To ensure consistency, we use the same spatial threshold for the Euclidean distance and the Fréchet distance, as their scales are directly comparable. The Fréchet distance reflects the shortest constraint (or "leash") needed for both trajectories to move along their paths simultaneously, capturing their geometric similarity over time.

After building the graph, we refine it by removing impossible connections and determining which tracklets should be fused, see Algorithm 2. We start by sorting the tracklets by the temporal length first and secondly by the x-coordinate of the first point in the tracklet, so that more weight is put on fusing longer tracklets which are more reliable. Only when the tracklets originate from the same same camera and when they have temporal overlap, we do not add them as candidates for fusion, in any other case, the tracklets can be fused. Lastly when selecting one of the candidates we first sort by the amount of temporal overlap and then by the distance to avoid fusing short tracklets with long ones first.

Finally we compute the arithmic mean between the fused tracklets. This tracking by curve matching pipeline improves upon the tracking by detection paradigm in three key ways:

- **More informative earlier temporal association:** Using bounding boxes in the image plane instead of points on the ground plane gives the trackers more information to associate detections such as the size and orientation of the broiler.
- **Projection correction:** Early temporal association makes it possible to correct the projection for broilers that are standing up versus sitting down.
- **More informative curve matching procedure:** Tracklets in the ground plane offer more information when comparing their Fréchet Distance instead of comparing points using Euclidean distance.

The parameters for these methods were tuned using Intel Xeon E5-2620 cpus while the inference performance was tested on a Ryzen 5 5600G Processor. Additional details on software and hardware will be provided in the github repository.

2.2.4. Performance Metrics

Each subtask employs metrics tailored to its objectives and challenges, as detailed in the following paragraphs.

Single-view detection: The performance of the detection model is evaluated using the average precision (AP) metric, a standard in object detection tasks (Zaidi et al., 2022). Specifically, we use $AP_{50\text{-}95}$, which calculates the mean AP across different intersection over union (IoU) thresholds ranging from 0.50 to 0.95 in increments of 0.05, providing a comprehensive measure of detection accuracy.

Multi-view detection: Multiple Object Detection Accuracy (MODA) and Multiple Object Detection Precision (MODP) are used as the primary performance metrics (Garofolo et al., 2006) and are the two main metrics used by scholars for evaluating multi-view detection algorithms (Hou et al., 2020; Hou and Zheng, 2021; López-Cifuentes et al., 2022; Vora et al., 2023). Additionally, we report precision and recall. In Appendix B we detail the difference in obtaining TPs for the image plane bounding box detections versus the multi-view ground plane detections. We evaluate this subtask seperately since its output can be used as is when tracking is not necessary for the specific use case.

Multi-view tracking: evaluates the model's capability to consistently track individual broilers over time. Commonly used tracking metrics are used and include IDF1 (Identity F1-score) (Ristani et al., 2016), which balances precision and recall at the identity level, Recall (Rcll), Precision (Prcn), Multiple Object Tracking Accuracy (MOTA), and Multiple Object Tracking Precision (MOTP) (Garofolo et al., 2006). Other metrics, such as Mostly Tracked (MT), Mostly Lost (ML), and ID-related statistics, including Identity Switches (IDs) and Fragmentations (FM) (Milan et al., 2016), are also used to capture tracking continuity, robustness, and identity preservation. Full metric explanations are available in Appendix B.

Both the multi-view detection and tracking metrics require a distance threshold for matching detections and tracks. This threshold should reflect the physical size of the animals which is needed for both ground plane localisation and tracking tasks. In scholarly work on multi-view pedestrian detection, a single threshold value of 0.5 meters is used as the approximated average width of a human body (Vora et al., 2023). During the 6-week lifespan, the broilers keep growing, and their average size increases, as shown in Figure 9. The coefficient values were empirically determined based on real-

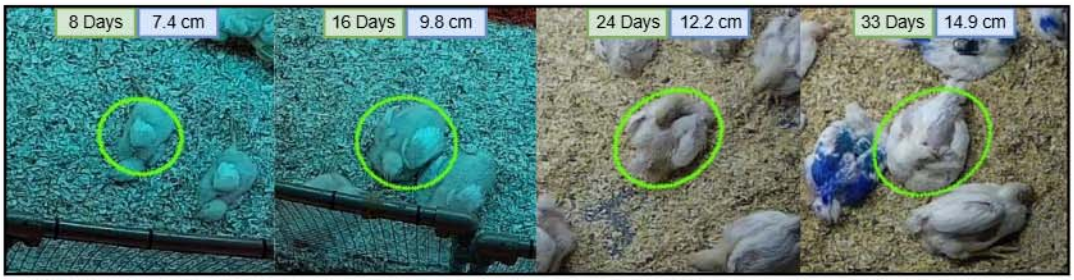

Figure 9: Visualization of the threshold for broilers at different ages. The age and radius are shown in the green and blue boxes respectively.

life measurements of broilers at different ages. This ensures that our metric calculation is consistent with the physical size of the broilers at different stages of their growth. We determine the threshold as a linear function of age, according to equation (1):

$$\mathrm{R} = 0.3\mathrm{cm} \times \mathrm{age} + 5\mathrm{cm} \tag{1}$$

3. Results and discussion

This section presents the experimental results of our multi-view, multi-object detection and tracking pipelines. These results all use per view bounding boxes as input. We first test various hyperparameter settings for the YOLO models, such as model size and image input resolution. Additionally the fine-tuned performance is compared to the pretrained performance of YOLOV11X. Next, the mutli-view detection results are presented, the role of individual camera perspectives and the fusion of their complementary information is investigated. Finally, the TBDM baseline approach is compared against our proposed TBCM method for multi-view tracking. This comparison highlights the improvements in tracking consistency and accuracy achieved by our approach, as well as its robustness across various broiler growth phases.

3.1. Single-view detection performance

Figure 10 displays the AP@50-95 of the fine-tuned YOLO models. YOLO represents a series of state-of-the-art pre-trained model architectures for object detection, available in various parameter configurations. Using the

single-view detection dataset, discussed in section 2.1.2, the pretrained YOLO models can be fine-tuned. The dataset was divided into a 95% portion for model development and a 5% validation portion for model selection. The development portion was used in a 5-fold cross-validation procedure to evaluate performance robustness, with confidence intervals reported. Each fold is split 80%|10%|10% for training, validation and testing respectively. The validation portion was exclusively used to select the optimal training epoch for the final model trained on the complete development set. For all dataset splits, samples were distributed across the entire time range of the experiment, ensuring temporal diversity in each partition.

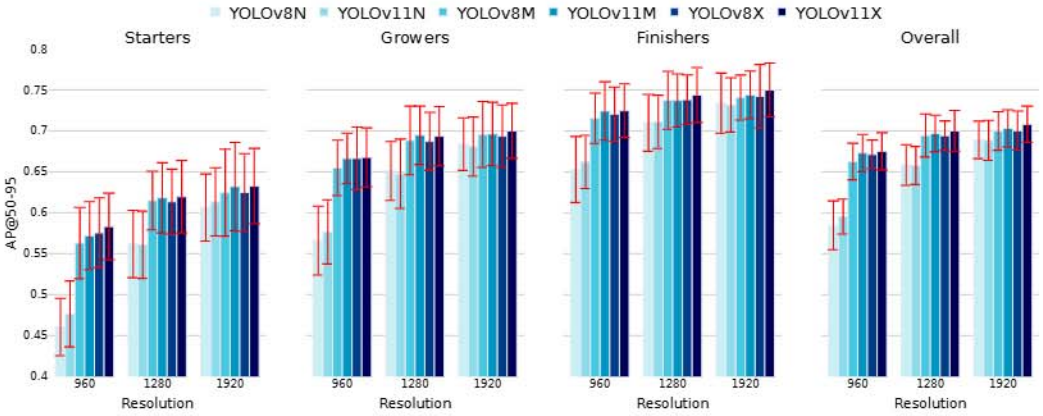

Figure 10: Single-view performance of different YOLO models (v8 and v11) at different input resolutions for the three dietary phases and overall with confidence intervals.

AP was assessed across input resolutions and YOLO model sizes for the three dietary phases. The significant impact of resolution on AP can be attributed to the presence of numerous small objects in the dataset, including very young broilers in the starter phase and animals positioned far from the camera. Subsequent results for multi-view detection and tracking are obtained using the YOLOV11X model with an input resolution of 1920x1080 pixels which achieves the highest overall AP@50-95 of 70.8%.

Figure 11 shows examples of incorrect detections for the single-view detection model. Both false positives and double detections will be propagated into the next tasks. Some of these errors will be filtered out by thresholding, but others will persist, emphasizing the need for further improvements in object detection. Missed and distant detections however can be adressed by combining multiple cameras. If a chicken is missed in one camera, another

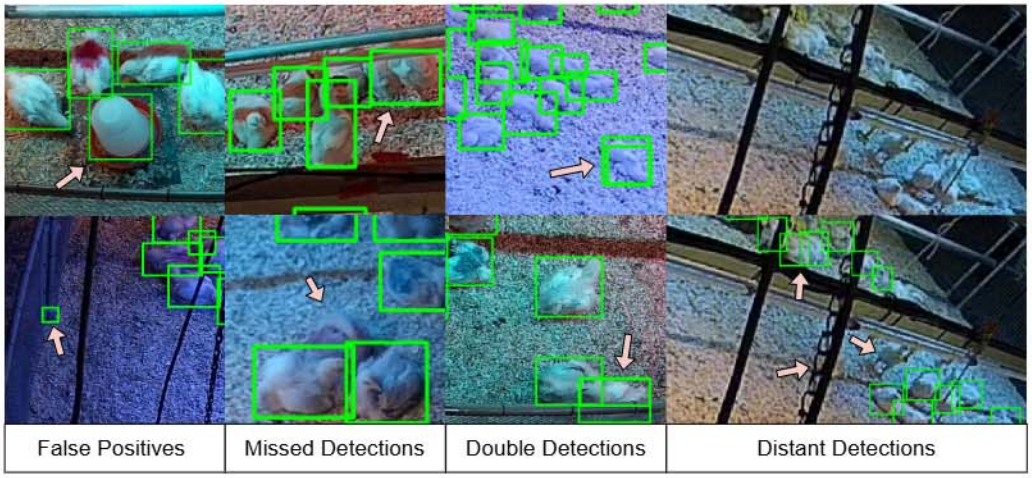

Figure 11: Examples of incorrect detections for the single-view detection model. The left three columns each include two examples of False Positives, Missed Detections and Double Detections. The right column shows an example of broilers far away from the camera. The green bounding boxes are obtained by the YOLO11X model.

camera could detect the chicken correctly. By having multiple cameras, the pen can have better coverage, allowing for the removal of these noisy detections.

Table 4: AP50-95 reported for the three dietary phases. Results were obtained on a finetuned YOLO11X model and on the pre-trained YOLO11X model, using the bird class as the target class. In both cases, an input resolution of 1920 x 1080 pixels was used.

Finetuned	Starter	Grower	Finisher	Overall
✗	1.58 ±1.5	11.16 ±3.0	21.8 ±1.6	13.94 ±1.3
✓	63.3 ±4.6	70.0 ±3.4	74.9 ±3.3	70.8 ±2.2

Table 4 also demonstrates the significant performance improvement after fine-tuning the YOLOV11X object detector on our dataset. Although the publicly available YOLOV11X version is pretrained on the COCO dataset (Lin et al., 2014) which does contain a bird class, it generalizes very poorly to our dataset, especially for younger broilers.

3.2. Multi-view detection performance

In Table 5, we report MODA, MODP, precision and recall on the multi-view detection dataset that was introduced in section 2.1.2. For tuning and

evaluating the multi-view detection pipeline, a 5-fold cross-validation procedure was used. Each fold was split 50%|50% for validation and testing respectively. For all dataset splits, samples were distributed across the entire time range of the experiment, ensuring temporal diversity in each partition.

Hyperparameter tuning was performed on the validation set of each fold after which we tested using the best parameters of each fold on the test set. However, we found that the best parameters for each fold were not significantly different from one fold to another. For this reason, we average the values of the best parameters of each fold to obtain the final parameters used for the multi-view detection pipeline, see Appendix D for hyperparameter details. In the table we include both the results with the same parameters averaged across folds and the results with unique parameters per fold.

Table 5: Multi-view detection results with 95% confidence intervals across folds for the different broiler phases. We report MODA, MODP, precision and recall. The first four rows show results with same parameters averaged across folds, while the last row shows results with unique parameters per fold. Overall† shows the result when using the tuned parameters for each fold.

Phase	MODA↑	MODP↑	Precision↑	Recall↑
Starter	84.9 ±2.8	63.9 ±1.2	92.0 ±1.6	93.0 ±1.2
Grower	85.9 ±0.9	65.6 ±0.5	93.1 ±0.5	92.8 ±0.5
Finisher	86.5 ±1.5	65.9 ±0.7	93.2 ±0.8	93.3 ±0.7
Overall	86.0 ±0.9	65.5 ±0.5	93.0 ±0.6	93.1 ±0.4
Overall†	85.8 ±1.1	65.4 ±1.6	92.9 ±0.7	93.0 ±0.7

The multi-view detection results, as shown in Table 5, demonstrate strong performance across the broiler growth phases, with the Finisher phase recording a MODA of 86.5% and precision of 93.2%, followed closely by the Grower phase at 85.9% MODA and 93.1% precision, and the Starter phase at 84.9% MODA and 92.0% precision. The overall performance across all phases remains robust, achieving a MODA of 86.0%, precision of 93.0%, and recall of 93.1%, reflecting the effectiveness of the detection system. Two key factors influence these results: the distance-based point selection step for approximating the broiler's location on the ground plane and the quality of the per-view detections. An analysis of different heuristics for feet approximation is detailed in Appendix C.

In the point selection process, the position of the broilers' feet is estimated

along the line between the bottom center and the center of the bounding box. The error in this estimation is constrained by the size of the bounding box, which increases with the size of the broilers. As broilers grow from the Starter to the Finisher phase, their larger bounding boxes lead to greater absolute errors when projecting their locations onto the ground plane. This amplifies the complexity of the fusion step, as the distance between projections of corresponding points from multiple views grows with these errors, potentially making it harder to match detections accurately across views.

Meanwhile, the quality of per-view detections, as discussed in 3.1, improves with broiler size. Larger broilers, such as those in the Finisher phase, are detected with greater accuracy compared to smaller ones in the Starter phase, due to their increased visibility and distinct features. This enhancement in detection quality provides more reliable inputs to the fusion process, potentially offsetting the challenges posed by larger position estimation errors. Across the phases, these two factors—position estimation error and detection quality—interact to shape the multi-view detection performance, as reflected in the results.

Figure 12 displays a sample from the test set with all four input views. The last row emphasizes the pipeline's capability to manage occlusions: despite multiple broilers being obscured by lighting equipment in camera 3, they are visible in camera 4, allowing the correct detection of the broilers. In Figure 13, a detailed analysis of the per camera performance is shown. The pen is divided into 30x30 cm squares, for each of the squares the MODA is calculated based on the single-view detections. For cameras 1 and 4, the entire pen is almost fully visible, which can be seen in Figure 12. However, performance significantly decreases for detections in the other half of the pen. This decline occurs because, for objects further away from the camera, small discrepancies in the selected point's position lead to large variations in the projected point on the ground plane.

Additionally, detecting broilers becomes more challenging as they move farther from the camera because they appear smaller in the images, making them occupy fewer pixels. Occlusions caused by the lighting equipment can also be seen in this figure; for cameras 1 and 4, there is a notable decrease in MODA in the fourth row from the bottom, while for cameras 3 and 4, the effect is less severe. Detections that are far away from the camera are removed as discussed in section 2.2.2, this removal threshold was fine-tuned on the validation set and closely aligns with the areas where performance begins to decline. The removal threshold value slightly exceeds the pen width (4

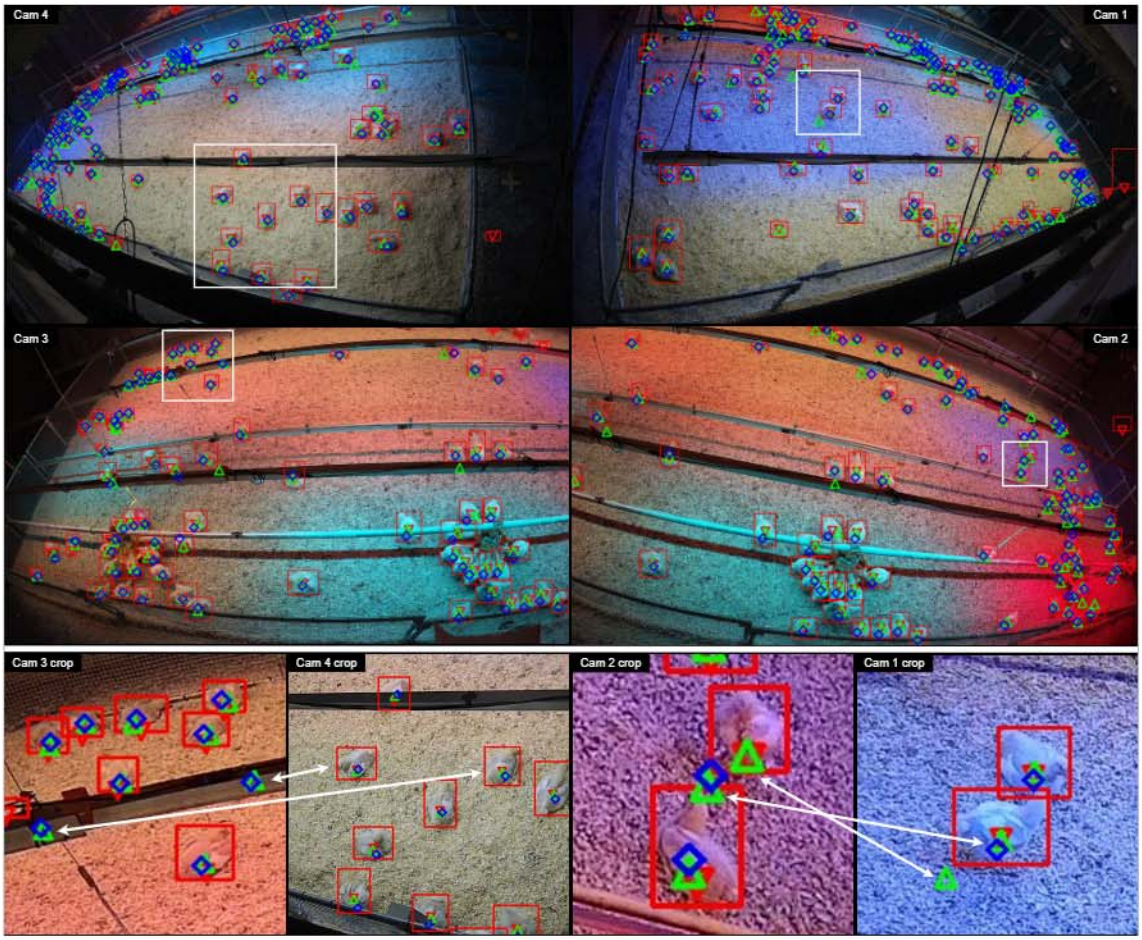

Figure 12: This figure shows one sample of the grower phase. The top two rows show all four input images with back-projected ground truths and predictions. In the left two crops of the last row we show the model's ability to handle occlusions due to the multi-camera setup. The right two crops show a double detection due to incorrect fusion of two detections. The ground truth labels are indicated with $\diamond$, the selected points using $\triangledown$ and the final predictions using $\triangle$. The bounding boxes $\square$ are the single view detection results. Figure is best viewed in color.

meters) at 440 cm. We can conclude that every camera adds complementary information, and all cameras are needed for the comprehensive detection of the whole pen. However, Figure 12 also shows a limitation of the pipeline in the two rightmost crops. Although the broilers are accurately detected in each individual camera, the projection on the ground plane is not correct because the broiler is standing up. As the distance between the projected points exceeds the set threshold, the detections fail to merge, leading to a false positive.

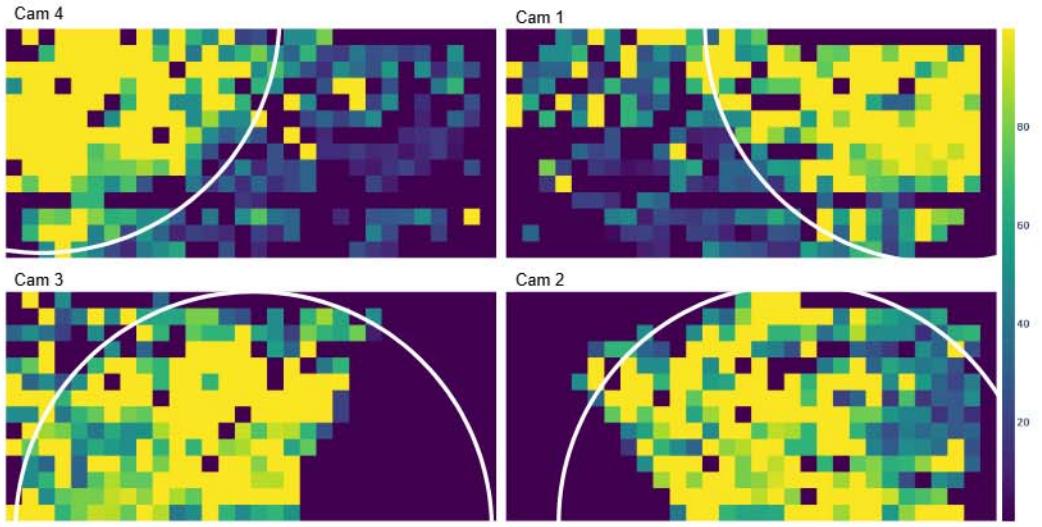

Figure 13: Performance of each individual camera. The pen is divided into 30x30 cm squares and shows the MODA for each square based on the single-view detections. The white circles indicate the remove threshold radius for each camera of which the center is each respective camera position. The figure shows that each camera adds complementary information.

3.3. Multi-view tracking performance

The ground plane tracking dataset, introduced in section 2.1.2, consists of six 30 second segments with on average 136 tracked broilers at 25 frames per second. Similar to how the the train and validation splits are made for the WildTrack dataset Chavdarova et al. (2018), each 30 second segment is split up into 10 second clips, totalling 18 samples. Again, a 5-fold cross-validation procedure is used, randomly splitting the samples 50%|50% validation and test respectively for each fold. Hyperparameters are tuned in each fold on the validation set. The multi-view tracking performance is presented in Table 6

Table 6: Comparison of tracking performance between the baseline tracking-by-detection-matching (TBDM) approach and our proposed tracking-by-curve-matching (TBCM) method. Each method is tested with both SORT and muSSP trackers (on ground plane for TBDM and in image plane for TBCM). Bold values indicate best average performance for each metric. Values are presented as mean ± 95% confidence interval.

	TBDM		TBCM	
Metric	SORT	muSSP	SORT	muSSP
MOTA↑	0.81 ± 0.00	0.81 ± 0.01	0.81 ± 0.00	0.81 ± 0.01
MOTP↓	3.96 ± 0.04	4.06 ± 0.04	**3.87** ± 0.06	3.89 ± 0.07
Rcll↑	0.88 ± 0.00	0.89 ± 0.00	0.89 ± 0.00	**0.90** ± 0.01
Prcn↑	**0.93** ± 0.00	**0.93** ± 0.00	0.92 ± 0.00	0.91 ± 0.01
MT↑	0.83 ± 0.01	0.83 ± 0.01	0.84 ± 0.01	**0.88** ± 0.01
PT↑	**0.13** ± 0.01	0.12 ± 0.00	0.10 ± 0.00	0.07 ± 0.00
ML↓	**0.05** ± 0.00	**0.05** ± 0.00	0.06 ± 0.00	0.06 ± 0.01
IDF1↑	0.85 ± 0.00	0.86 ± 0.01	**0.89** ± 0.00	**0.89** ± 0.01
FM↓	2111 ± 240	2029 ± 144	1276.0 ± 58.6	**1045** ± 166
IDs↓	1077 ± 164	929.4 ± 66.6	395.6 ± 11.5	**377.6** ± 91.9
IDt↓	372 ± 117	458.4 ± 25.4	97.8 ± 10.8	**73.8** ± 8.2
IDa↓	367 ± 159	242.0 ± 64.4	205.4 ± 13.4	**199.8** ± 45.7
IDm↓	30.8 ± 6.5	37.8 ± 2.3	19.4 ± 4.0	**12.6** ± 3.4

where the average performance across the five folds with the 95% confidence intervals is reported.

Compared to the baseline approach TBDM, our method TBCM achieves superior tracking consistency across both tested variants (SORT and muSSP). The TBCM approach with muSSP achieves the highest IDF1 score of 89%, showing an improvement over the TBDM baseline (86% with muSSP). This enhanced performance is further evidenced by significant reductions in identity switches (IDs) and identity transfers (IDt). TBCM reduces IDs by more than half compared to TBDM and reduces fragmentations (FM) by nearly 50%, with the lowest fragmentation count (1045) achieved by TBCM with muSSP. While TBDM maintains a slightly higher precision (93%), TBCM with muSSP achieves the best recall (90%) and notably improves the proportion of mostly tracked objects (MT) to 88%. While MOTA values remain comparable across all methods, the best MOTP score is achieved by TBCM (SORT) with 3.87cm, indicating improved localization accuracy.

Figure 14 shows a qualitative comparison between the TBDM and TBCM

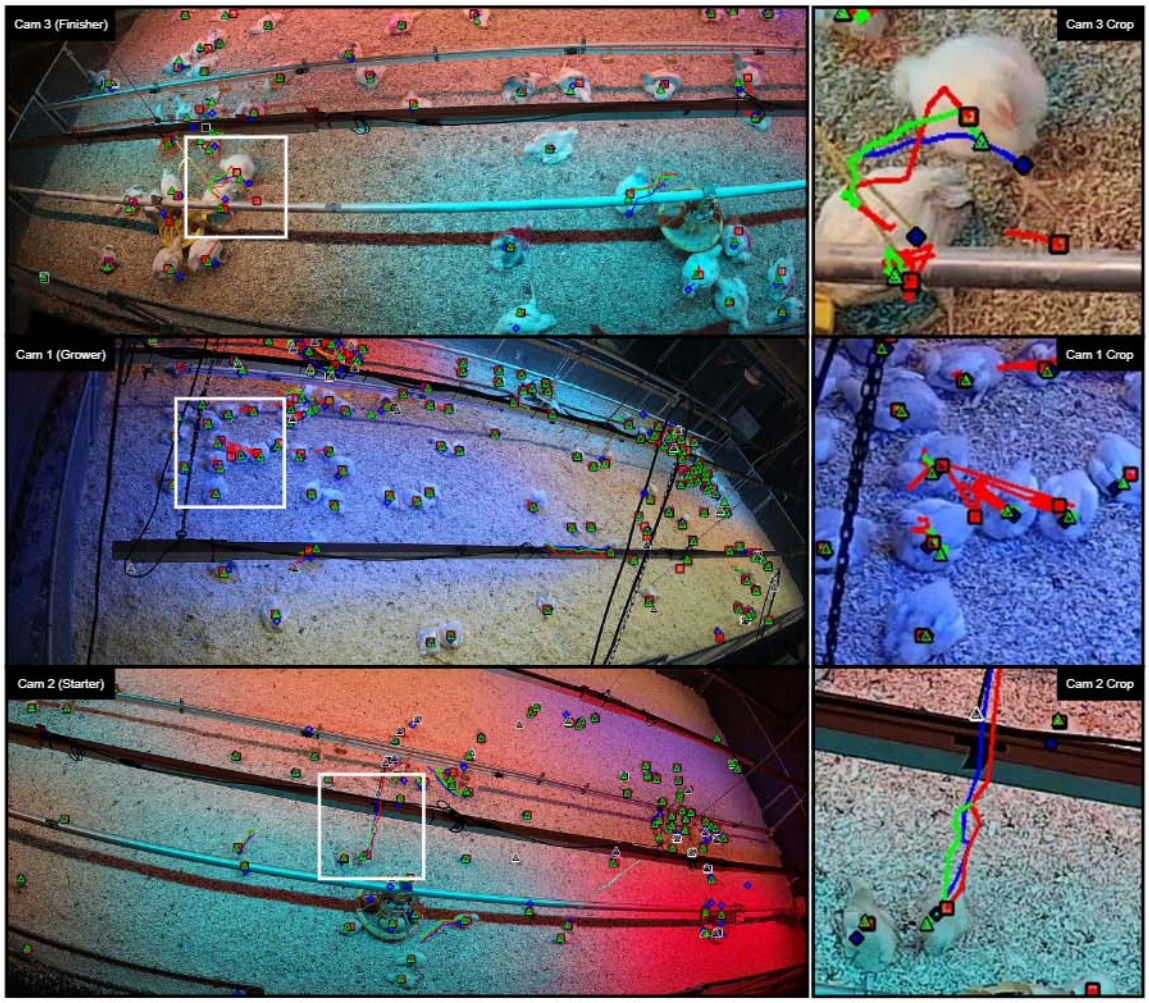

Figure 14: This figure compares the tracking methods TBDM and TBCM (muSSP) where ground truth labels are indicated with ◇, TBDM (muSSP) predicted tracks with □ and TBCM (muSSP) with △. Black and white tracks indicate lost tracks. To keep the figure clear, we only show the last four seconds (100 frames) of each track. The top row shows the improved fusion due the correction for moving broilers. The second row shows the degraded performance of TBDM due to the close proximity of the the broilers. The last row shows TBCM's inability to connect detections for longer temporal gaps. Figure is best viewed in color.

Table 7: Tracking performance across different broiler growth phases using TBCM (muSSP). Values are presented as mean ± 95% confidence interval.

Metric	starter	grower	finisher	OVERALL
MOTA↑	0.78 ± 0.02	0.82 ± 0.02	0.85 ± 0.01	0.81 ± 0.01
MOTP↓	2.75 ± 0.05	4.06 ± 0.11	4.82 ± 0.15	3.89 ± 0.07
Rcll↑	0.88 ± 0.01	0.90 ± 0.01	0.93 ± 0.01	0.90 ± 0.01
Prcn↑	0.90 ± 0.01	0.91 ± 0.01	0.92 ± 0.01	0.91 ± 0.01
MT↑	0.84 ± 0.01	0.87 ± 0.02	0.91 ± 0.01	0.88 ± 0.01
PT↑	0.09 ± 0.00	0.08 ± 0.01	0.04 ± 0.01	0.07 ± 0.00
ML↓	0.07 ± 0.01	0.05 ± 0.01	0.05 ± 0.01	0.06 ± 0.01
IDF1↑	0.86 ± 0.01	0.89 ± 0.01	0.91 ± 0.01	0.89 ± 0.01
FM↓	325.40 ± 39.16	428.60 ± 107.22	291.00 ± 31.82	1045.00 ± 166.28
IDs↓	119.00 ± 18.44	153.20 ± 43.53	105.40 ± 33.50	377.60 ± 91.91
IDt↓	30.00 ± 6.76	26.00 ± 4.68	17.80 ± 4.04	73.80 ± 8.23
IDa↓	70.00 ± 12.11	79.00 ± 20.16	50.80 ± 14.37	199.80 ± 45.70
IDm↓	5.20 ± 2.00	4.60 ± 1.18	2.80 ± 0.96	12.60 ± 3.37

methods. The figure highlights the benefits of the TBCM approach as well as a limitation compared to TBDM. In the previous section, we discussed the limitation of the multi-view detection method, where the incorrect fusion of detections from different broilers is caused by the inacurate projection of the broiler's location on the ground plane. This problem is propagated in the tracks produced by the TBDM algorithm. By adjusting the projection for broilers that are moving around, our TBCM approach can adress this problem and fuses the two tracklets correctly. As shown in Table 6, metrics related to track identities, such as IDF1, IDs and FM are greatly improved with the TBCM method, an example of why is shown on the second row in the qualitative figure. The TBDM approach wrongly accociates detections from different broilers, creating inconsistent tracks especially when broilers are close to each other. Due to early association of the bounding boxes instead of points on the ground plane, our tracklet fusion method can utilize more information to determine which tracklets should be fused. A limitation of the TBCM method, however, is illustrated on the third row. TBCM does not take into account motion information for the fusion of tracklets when there are no overlapping timesteps between the tracklets. This makes the fusion of these tracklets much harder, resulting in more fragmentations.

Table 7 presents the performance metrics for the various dietary phases for the TBCM method that uses muSSP for obtaining per-view tracklets.

The results indicate that tracking performance improves across the broiler growth phases, with the finisher phase achieving the highest scores in metrics like IDF1 (91.0%) and MOTA (85.0%). The starter phase shows the lowest performance, with metrics such as MOTA (78.0%) and higher fragmentation (FM: 325.40). As discussed in 1, younger broilers are generally more active, which makes tracking more challenging. However, for younger broilers, MOTP tends to be lower (2.75), due to their smaller bounding boxes, which allow for more accurate projections.

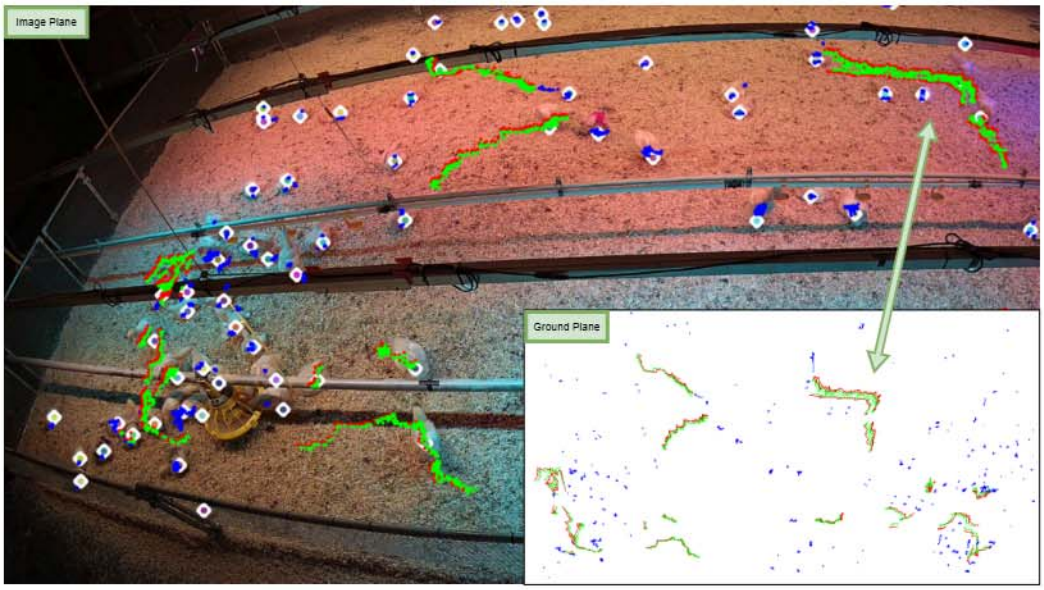

Figure 15: Visualization of what the algorithm considers standing and walking broilers. All curves shown are the filtered curves (distant ones removed) from each camera, shown both in the image plane and ground plane. The curves represent 10 seconds of video. Blue lines represent standing broilers and red/green lines represent walking broilers. The green lines represent the corrected red curves for the broilers that are considered walking from different cameras. The green arrow highlights a walking broiler which is tracked in the image plane by three seperate cameras (three red lines and three green lines), due to the correction, the green lines overlap much more with each other, allowing for correct curve matching. Diamond $\diamond$ markers (with a color for each track id) represent the ground truth locations at the timestep shown.)

Figure 15 shows a visualization of what the algorithm considers standing and walking broilers and how the algorithm corrects for the walking broilers allowing for correct curve matching. Within the tracking window a broiler can only be considered as either standing or walking, this is a limitation of

the algorithm.

Method	Process	5s	15s	30s	Linear
TBDM (muSSP)	per frame multi-view detection	8.2 ± 0.2	24.7 ± 1.0	49.4 ± 2.8	✓
	muSSP detection matching	39.8 ± 1.2	138.7 ± 5.4	334.6 ± 23.4	✗
	Total	48.0 ± 0.8	163.4 ± 3.7	384.0 ± 15.2	✗
TBCM (muSSP)	muSSP 2D tracking (4 views)	92.7 ± 4.9	315.3 ± 30.5	692.4 ± 102.6	✗
	tracklets to curves	2.8 ± 0.1	8.8 ± 0.5	17.8 ± 1.4	✓
	curve matching	2.5 ± 0.3	18.4 ± 4.3	64.2 ± 23.4	✗
	Total	98.0 ± 2.8	342.5 ± 17.1	774.4 ± 55.5	✗
TBDM (SORT)	per frame multi-view detection	8.4 ± 0.2	25.0 ± 1.0	50.5 ± 2.9	✓
	SORT detection matching	1.3 ± 0.0	3.8 ± 0.1	7.6 ± 0.3	✓
	Total	9.7 ± 0.1	28.8 ± 0.7	58.0 ± 1.9	✓
TBCM (SORT)	SORT 2D tracking (4 views)	3.5 ± 0.1	10.5 ± 0.6	21.0 ± 1.7	✓
	tracklets to curves	2.8 ± 0.1	8.6 ± 0.5	17.4 ± 1.4	✓
	curve matching	2.5 ± 0.3	16.7 ± 3.9	55.9 ± 19.5	✗
	Total	8.8 ± 0.2	35.8 ± 2.2	94.4 ± 10.4	✗

Table 8: Processing times (in seconds) for different tracking methods on an AMD Ryzen 5 5600G Processor. The columns indicate the duration of video processed at 25 fps. Note that YOLOv11x detection times are not included as they are well-documented online. The Linear column indicates whether the process scales linearly (✓) or non-linearly (✗) with video duration.

Table 8 presents the processing times for different tracking methods on an AMD Ryzen 5 5600G Processor, with the column names indicating the video durations processed at 25 fps. Note that YOLOv11x detection times are not included as they are required for all methods. While the 2D tracking for TBCM could easily be parallelized for each view, the timings here are reported for the sequential case. Furthermore while muSSP is implemented in C++ for matching detections faster, our method is implemented in Python. While the per frame multi-view detection, converting tracklets to curves (projecting them onto the ground plane) and SORT tracking/detection matching are all linear, the curve matching and muSSP detection matching steps are non-linear since they are global operations that scale with the number of detections. While the TBDM (SORT) method is the fastest method it has the lowest tracking performance.

4. Conclusion

In this study, we proposed a multi-camera, individual broiler detection and tracking method designed to localize and track broilers in experimental pen setups. Combining a state-of-the-art object detection model, a dynamic point selection strategy, and precise geometric fusion, we achieved an average MODA of 85.8% and 93% recall for broiler localisation, showcasing the pipeline's robustness across different broiler ages. Our novel tracklet fusion approach improves the tracking-by-detection paradigm by reliably tracking up to 5% more broilers. By leveraging complementary viewpoints from multiple cameras, our system can better handle occluded broilers, significantly improving detection and tracking performance compared to single-camera approaches. Broiler tracks can provide useful insights into behaviour. As illustrated in Figure 1, these tracks can be post-processed to identify different behavioral patterns, such as activity levels or resource usage. This demonstrates the potential to automate continuous behavioral analysis in research settings, significantly reducing the need for manual observation while enhancing objectivity and consistency. We also open source the comprehensive labeled dataset that includes broilers at varying ages, from 5 day to 38 days old, providing a valuable resource for the community to advance research in this domain. A notable strength of our system is its unsupervised nature where only the object detection component requires labeled training data. This is a deliberate choice given the labor-intensive nature of annotating full pens of broilers and highlights the scalability of the approach for larger pen sizes. Despite these advantages, challenges remain. Poor visibility in corner areas impacts performance, this issue could be addressed through optimized camera placement or by adding additional cameras beyond our current setup of one camera per $9m^2$. The system's flexibility allows users to determine the extent of the pen surface to be monitored, providing adaptability to specific requirements of different setups. In future work we aim to extend the system's utility beyond research settings and test a multi-camera system in commercial broiler farms as our method is able to improve tracking results particularly when broilers are closely grouped together. Additionally, we plan to investigate post-processing techniques to extract behavioral features from the tracks. Lastly, we intend to improve our tracklet fusion approach by incorporating motion features during the fusion process to further improve tracking results.

5. Acknowledgments

The imec.icon project WISH is a research project bringing together academic researchers and industry partners. Project WISH is co-financed by imec and receives financial support from Flanders Innovation & Entrepreneurship (project nr. HBC.2021.0664).

We gratefully acknowledge the support of ILVO Vlaanderen, who organized the experimental setup, took care of the broilers, and provided us with domain knowledge.

Sam Leroux and Pieter Simoens acknowledge the financial support of the Flanders AI Research (FAIR) program.

Appendix A. Camera calibration

Camera calibration involves estimating the intrinsic and extrinsic parameters of the camera. Intrinsic parameters are properties of the camera itself of which the value is invariant to the scene being captured. Extrinsic parameters, on the other hand, describe the position and orientation of the camera with respect to the world coordinate system. In our case, the world coordinate system has its origin in one of the pen corners and its coordinate axes aligned with the fences of the pen and the pen floor as XY-plane. Accurate calibration is needed to map coordinates in the 2D image frame of one camera onto a position in the world coordinate system, or to map 2D coordinates in the image frame of one camera into image frame coordinates of another camera. The calibration procedure utilized for the WildTrack dataset (Chavdarova et al., 2018) is closely followed.

Intrinsics: To perform the intrinsic calibration of each camera, the procedure detailed in OpenCV (2023) employing a Charuco board is used. To detect the Charuco board, the Aruco library can be used which is part of OpenCV (Bradski, 2000). The detected corner points can then be used to obtain the camera matrix and distortion coefficients. The camera matrix is a 3×3 matrix that includes the camera's focal length and principal point. The distortion coefficients determine the radial and tangential distortion of the camera optics.

Extrinsics: The process of extrinsic calibration is fundamental for establishing a robust and accurate relationship between the camera coordinate system and the world coordinate system. This calibration is performed manually, since self-calibration techniques rely on distinct visual markers that are

absent in broilers (Xu et al., 2021; Tang et al., 2019). 43 strategically chosen points are marked within the pen, manually measuring their 3D coordinates within the world coordinate system. Simultaneously, the corresponding 2D positions of these points within the camera views are recorded. Using the acquired 3D-2D correspondences, the rotation and translation parameters of each individual camera pose can be computed by minimizing the reprojection error. This can be achieved using the OpenCV method solvePnP that uses the non-linear Levenberg-Marquardt minimization scheme (Madsen et al., 2004).

Bundle Adjustement: Finally, joint optimization of the calibration parameters is performed for the four cameras through a technique known as bundle adjustment. Through a set of images capturing 3D points from different perspectives, bundle adjustment can be characterized as the task of refining the 3D coordinates representing the scene geometry concurrently with the intrinsic and extrinsic camera parameters. This refinement is carried out based on an optimality criterion that considers the corresponding image projections of all points. An additional set of 699 2D-2D correspondences is labelled across the four camera viewpoints to facilitate this optimisation process. These points correspond to a total of 238 unique points in the world coordinate system. To solve this nonlinear least-squares problem, the SciPy Python library (Virtanen et al., 2020) is used.

Calibration accuracy Figure A.16 highlights the quality of the mutli-camera calibration. The bundle adjustment step greatly improves the joint camera calibration and is crucial to the performance of the pipeline.

Appendix B. Additional metric details

In the following paragraphs we detail the difference in obtaining TPs for the single-view detections versus the multi-view detections.

AP@50-95: Predicted bounding boxes are considered as a positive detection of a ground truth bounding box when the Intersection-over-Union (IoU) exceeds a predefined threshold. The matched prediction with the highest confidence is considered as a TP, and all others as False Positives (FP). The number of False Negatives (FN) is determined as the number of ground truth bounding boxes for which none of the predictions has an IoU larger than the threshold. **Multi-View Metrics:** To match ground plane detections to their respective ground truth, minimum cost assignment between ground truth objects and predictions is used based on a predefined distance

Table B.9: Overview of evaluation metrics used in this work.

Subtask	Metric	Description
Single-view detection	AP50-95	Measures how well the detector finds broilers in single camera views. The score averages detection accuracy across different overlap thresholds (50% to 95%) between predicted and actual broiler locations
Multi-view detection	MODA	Overall detection quality score that penalizes both missing broilers (false negatives) and detecting non-existent broilers (false positives)
	MODP	Measures how accurately the detected broiler positions match their true positions on the ground plane
	Precision	Out of all broilers the system detected, what percentage were actually real broilers (rather than false detections)
	Recall	Out of all real broilers present, what percentage did the system successfully detect
Multi-view tracking	MOTA	Overall tracking quality score that considers missed detections, false detections, and cases where broiler identities get mixed up over time
	MOTP	How precisely the system tracks broiler positions - measured as the average distance between predicted and actual broiler locations
	MT	Percentage of broilers that were successfully tracked for more than 80% of the ground truth trajectory
	ML	Percentage of broilers that were rarely or never tracked (less than 20% of the ground truth trajectory)
	IDF1	How well the system maintains correct broiler identities throughout tracking, considering both detection accuracy and identity consistency
	FM	Number of times the system temporarily loses track of a broiler before finding it again
	IDs	(ID switch) Number of times the system incorrectly assigns a new identity to a broiler that has already been tracked, mistakenly treating it as a different broiler.
	IDt	(ID transfer) Number of times an existing track is reassigned to a different broiler, leading to an incorrect continuation of identity.
	IDa	(ID ascension) The number of times a new identity is assigned to a broiler that was previously untracked.
	IDm	(ID migration) The number of times a broiler that has already been tracked is assigned a different identity, leading to an inconsistency in tracking.
	Precision	Out of all tracked broiler positions, what percentage correctly matched real broiler positions
	Recall	Out of all real broiler positions over time, what percentage were successfully detected

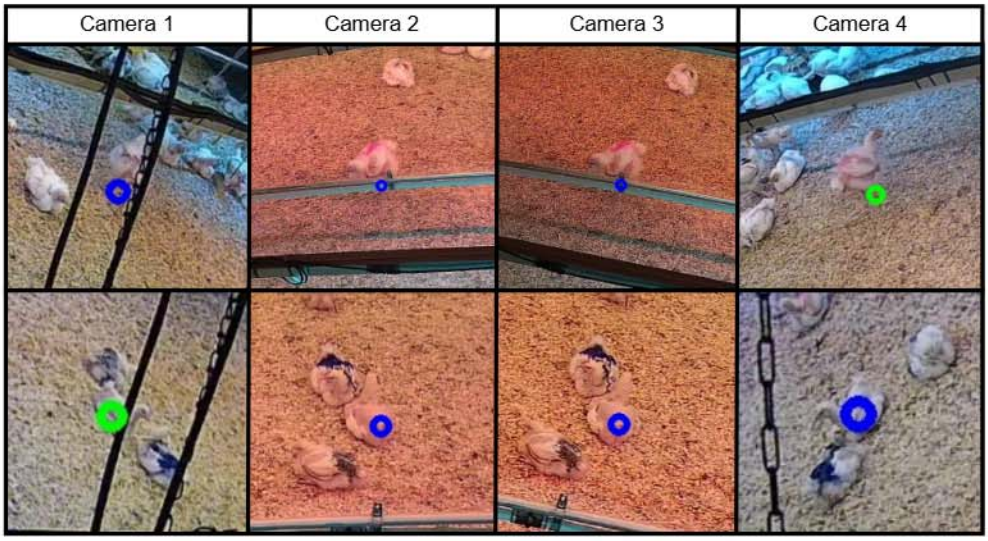

Figure A.16: Each row corresponds to a set of 4 time-synchronized frames as captured by the different cameras. To demonstrate the accuracy of our camera calibration, the chicken positions (green circles) are manually indicated in one view. These points are then first projected onto the ground plane and subsequently projected back into the other three viewpoints (blue circles).

threshold. The matched detections are considered TPs, and all others as FPs.

For the multi-view detection metrics, the code of Hou et al. (2020) is used while the py-motmetrics library Heindl et al. (2017) is used for the tracking metrics. Table B.9 provides an overview of the metrics used in this work.

Appendix C. Point selection strategies

To investigate the localisation errors introduced by the point selection and projection steps, we compare the performance of the distance-based point selection mechanism with the two other heuristic selection strategies introduced in section 2.2.2. We also experimented with the effect of not projecting points too distant from the camera. The results are reported in Table C.10.

From the results, we can conclude that both the distance-based point selection strategy and the distance-based removal strategy positively impact the detection performance. Not projecting bounding boxes of too-distant detections does not improve recall but does improve accuracy, which shows that only incorrect detections are removed.

Table C.10: Performance comparison of different point selection strategies and the addition of discarding distant points. The selection row refers to which point in the bounding box is selected for projection to a ground position. The removal row refers to whether too-distant points are discarded or not.

Point selection	Remove Threshold	MODA	MODP	Precision	Recall
center	✓	64.1 ±0.8	55.1 ±0.7	81.2 ±0.4	83.4 ±0.7
bottom	✓	76.7 ±2.3	51.2 ±0.5	87.0 ±1.2	90.4 ±1.0
distance	✗	76.9 ±1.7	56.1 ±0.4	86.5 ±1.0	91.3 ±0.6
distance	✓	86.0 ±0.9	65.5 ±0.5	93.0 ±0.6	93.1 ±0.4

Appendix D. Hyperparameters

Table D.11 details all the tuned hyperparameters for the methods discussed in the paper. 3D and 2D refer to the ground plane and image plane for the muSSP tracking method. The broiler radius (BR) depends on the age of the broiler, see section 2.2.4. Other hyperparameters could be tuned for the muSSP method, such as weight_source_sink, sigma_box_size and sigma_motion, but we found that the default values worked well for the current experimental setup. For more details on the hyperparameters, we refer to our github repository.

References

Aldridge, D., Owens, C., Maynard, C., Kidd, M., Scanes, C., 2022. Impact of light intensity or choice of intensity on broiler performance and behavior. Journal of Applied Poultry Research 31, 100216. URL: https://www.sciencedirect.com/science/article/pii/S1056617121000799, doi:https://doi.org/10.1016/j.japr.2021.100216.

Aviagen, 2022. Ross 308 broiler: Performance objectives. URL: https://aviagen.com/assets/Tech_Center/Ross_Broiler/RossxRoss308-BroilerPerformanceObjectives2022-EN.pdf.

Aydin, A., 2017. Development of an early detection system for lameness of broilers using computer vision. Computers and Electronics in Agriculture 136, 140–146. URL: https://www.sciencedirect.

Table D.11: Hyperparameters used for each tracking method across all experiments.

Component	Parameter	Multi-view detection	TBD		TBCM	
			SORT	muSSP	SORT	muSSP)
BB Detection	BB_th	0.51	0.51	0.51	0.51	0.51
Point Fusion	dist_min	17	17	17	17	17
	dist_max	449	449	449	449	449
	rem_th	408	408	408	-	-
Tracklet Fusion	avg_rem_th	-	-	-	380	380
	length_th	-	-	-	10	10
	R	-	-	-	0.8*BR	0.8*BR
	shift	-	-	-	0.075*BR	0.075*BR
	max_gap	-	-	-	30	30
SORT	min_hits	3	-	-	3	-
	max_age	25	-	-	15	-
	IoU_th	0.7	-	-	0.7	-
muSSP	sigma_dist	-	-	11	-	11
	max_jump_pass1	-	-	6	-	8
	max_jump_pass2	-	-	15	-	10
	split_length	-	-	18	-	20
	th_length	-	-	14	-	8
	max_dist	-	-	1.5 * BR	-	23

com/science/article/pii/S0168169917302296, doi:https://doi.org/10.1016/j.compag.2017.02.019.

Baxter, M., Bailie, C., O'Connell, N., 2018. An evaluation of potential dustbathing substrates for commercial broiler chickens. Animal 12, 1933–1941. URL: https://www.sciencedirect.com/science/article/pii/S1751731117003408, doi:https://doi.org/10.1017/S1751731117003408.

Baxter, M., Bailie, C., O'Connell, N., 2019. Play behaviour, fear responses and activity levels in commercial broiler chickens provided with preferred environmental enrichments. Animal 13, 171–179. URL: https://www.sciencedirect.com/science/article/pii/S1751731118001118, doi:https://doi.org/10.1017/S1751731118001118.

Bewley, A., Ge, Z., Ott, L., Ramos, F.T., Upcroft, B., 2016. Simple online and realtime tracking, in: 2016 IEEE International Conference on Image Processing, ICIP 2016, Phoenix, AZ, USA, September 25-28, 2016, IEEE. pp. 3464–3468. URL: https://doi.org/10.1109/ICIP.2016.7533003, doi:10.1109/ICIP.2016.7533003.

Bizeray, D., Leterrier, C., Constantin, P., Picard, M., Faure, J.M., 2000. Early locomotor behaviour in genetic stocks of chickens with different growth rates. Applied Animal Behaviour Science 68, 231–242. doi:10.1016/S0168-1591(00)00097-7.

Bradski, G., 2000. The OpenCV Library. Dr. Dobb's Journal of Software Tools .

Campbell, M., Miller, P., Díaz-Chito, K., Hong, X., McLaughlin, N., Parvinzamir, F., Martínez Del Rincón, J., O'Connell, N., 2024. A computer vision approach to monitor activity in commercial broiler chickens using trajectory-based clustering analysis. Computers and Electronics in Agriculture 217, 108591. URL: https://www.sciencedirect.com/science/article/pii/S0168169923009791, doi:https://doi.org/10.1016/j.compag.2023.108591.

Cao, L., Xiao, Z., Liao, X., Yao, Y., Wu, K., Mu, J., Li, J., Pu, H., 2021. Automated chicken counting in surveillance camera environments based on the point supervision algorithm: Lc-densefcn. Agriculture

11. URL: https://www.mdpi.com/2077-0472/11/6/493, doi:10.3390/agriculture11060493.

Cartucho, J., 2018. Openlabeling. https://github.com/Cartucho/OpenLabeling.

Cartucho, J., Ventura, R., Veloso, M., 2018. Robust object recognition through symbiotic deep learning in mobile robots, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2336–2341.

Chavdarova, T., Baqué, P., Bouquet, S., Maksai, A., Jose, C., Bagautdinov, T.M., Lettry, L., Fua, P., Gool, L.V., Fleuret, F., 2018. WILDTRACK: A multi-camera HD dataset for dense unscripted pedestrian detection, in: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018, Computer Vision Foundation / IEEE Computer Society. pp. 5030–5039. URL: http://openaccess.thecvf.com/content_cvpr_2018/html/Chavdarova_WILDTRACK_A_Multi-Camera_CVPR_2018_paper.html, doi:10.1109/CVPR.2018.00528.

CVAT.ai, 2024. Computer vision annotation tool (cvat). URL: https://doi.org/10.5281/zenodo.12771595, doi:10.5281/zenodo.12771595.

Engilberge, M., Liu, W., Fua, P., 2023. Multi-view tracking using weakly supervised human motion prediction, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1582–1592.

Fernandes, B.C.d.S., Martins, M.R.F.B., Mendes, A.A., Paz, I.C.d.L.A., Komiyama, C.M., Milbradt, E.L., Martins, B.B., 2012. Locomotion problems of broiler chickens and its relationship with the gait score. Revista Brasileira de Zootecnia 41, 2333–2336. URL: http://www.sbz.org.br.

Ferryman, J.M., Shahrokni, A., 2009. Pets2009: Dataset and challenge. 2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance , 1–6URL: https://api.semanticscholar.org/CorpusID:34646357.

Garofolo, J.S., Bowers, R., Moellman, D.E., Kasturi, R., Goldgof, D., Soundararajan, P., 2006. Performance evaluation protocol for face, person

and vehicle detection & tracking in video analysis and content extraction (vace-ii) clear - classification of events, activities and relationships. URL: https://api.semanticscholar.org/CorpusID:130025156.

Guo, Y., Aggrey, S.E., Oladeinde, A., Johnson, J., Zock, G., Chai, L., 2021. A machine vision-based method optimized for restoring broiler chicken images occluded by feeding and drinking equipment. Animals 11. URL: https://www.mdpi.com/2076-2615/11/1/123, doi:10.3390/ani11010123.

Guo, Y., Chai, L., Aggrey, S.E., Oladeinde, A., Johnson, J., Zock, G., 2020. A machine vision-based method for monitoring broiler chicken floor distribution. Sensors 20, 3179. URL: https://doi.org/10.3390/s20113179, doi:10.3390/s20113179.

Han, X., You, Q., Wang, C., Zhang, Z., Chu, P., Hu, H., Wang, J., Liu, Z., 2023. Mmptrack: Large-scale densely annotated multi-camera multiple people tracking benchmark, pp. 4849–4858. doi:10.1109/WACV56688.2023.00484.

Heindl, C., Toka, Valmadre, J., 2017. py-motmetrics [software]. https://github.com/cheind/py-motmetrics.

Hou, Y., Zheng, L., 2021. Multiview detection with shadow transformer (and view-coherent data augmentation), in: Shen, H.T., Zhuang, Y., Smith, J.R., Yang, Y., César, P., Metze, F., Prabhakaran, B. (Eds.), MM '21: ACM Multimedia Conference, Virtual Event, China, October 20 - 24, 2021, ACM. pp. 1673–1682. URL: https://doi.org/10.1145/3474085.3475310, doi:10.1145/3474085.3475310.

Hou, Y., Zheng, L., Gould, S., 2020. Multiview detection with feature perspective transformation, in: Vedaldi, A., Bischof, H., Brox, T., Frahm, J. (Eds.), Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part VII, Springer. pp. 1–18. URL: https://doi.org/10.1007/978-3-030-58571-6_1, doi:10.1007/978-3-030-58571-6_1.

Hwang, J., Benz, P., Kim, T., 2022. Booster-shot: Boosting stacked homography transformations for multiview pedestrian detection with attention.

CoRR abs/2208.09211. URL: https://doi.org/10.48550/arXiv.2208.09211, doi:10.48550/arXiv.2208.09211, arXiv:2208.09211.

ILVO, imec, 2022-2024. Wish research project. URL: https://ilvo.vlaanderen.be/en/research-projects/welzijnsbevorderende-monitoringsoplossingen-voor-kippen.

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., Michael, K., TaoXie, Fang, J., imyhxy, Lorna, Yifu, Z., Wong, C., V, A., Montes, D., Wang, Z., Fati, C., Nadar, J., Laughing, UnglvKitDe, Sonck, V., tkianai, yxNONG, Skalski, P., Hogan, A., Nair, D., Strobel, M., Jain, M., 2022. ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation. URL: https://doi.org/10.5281/zenodo.7347926, doi:10.5281/zenodo.7347926.

Li, G., Zhao, Y., Purswell, J.L., Du, Q., Chesser, G.D., Lowe, J.W., 2020a. Analysis of feeding and drinking behaviors of group-reared broilers via image processing. Computers and Electronics in Agriculture 175, 105596. URL: https://www.sciencedirect.com/science/article/pii/S0168169919305745, doi:https://doi.org/10.1016/j.compag.2020.105596.

Li, N., Ren, Z., Li, D., Zeng, L., 2020b. Review: Automated techniques for monitoring the behaviour and welfare of broilers and laying hens: towards the goal of precision livestock farming. Animal 14, 617–625. URL: https://www.sciencedirect.com/science/article/pii/S1751731119002155, doi:https://doi.org/10.1017/S1751731119002155.

Li, X., Zhao, Z., Wu, J., Huang, Y., Wen, J., Sun, S., Xie, H., Sun, J., Gao, Y., 2022. Y-BGD: broiler counting based on multi-object tracking. Comput. Electron. Agric. 202, 107347. URL: https://doi.org/10.1016/j.compag.2022.107347, doi:10.1016/j.compag.2022.107347.

Lima, J., Roberto, R., Silva Figueiredo, L., Simoes, F., Thomas, D., Uchiyama, H., Teichrieb, V., 2022. 3d pedestrian localization using multiple cameras: a generalizable approach. Machine Vision and Applications 33. doi:10.1007/s00138-022-01323-9.

Lin, T., Maire, M., Belongie, S.J., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: common objects in context, in: Fleet, D.J., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V, Springer. pp. 740–755. URL: https://doi.org/10.1007/978-3-319-10602-1_48, doi:10.1007/978-3-319-10602-1_48.

López-Cifuentes, A., Escudero-Viñolo, M., Bescós, J., Carballeira, P., 2022. Semantic-driven multi-camera pedestrian detection. Knowl. Inf. Syst. 64, 1211–1237. URL: https://doi.org/10.1007/s10115-022-01673-w, doi:10.1007/s10115-022-01673-w.

Madsen, K., Nielsen, H., Tingleff, O., 2004. Methods for non-linear least squares problems (2nd ed.) , 60.

McLaughlin, N., Rincon, J.M.D., Miller, P., 2015. Enhancing linear programming with motion modeling for multi-target tracking, in: 2015 IEEE Winter Conference on Applications of Computer Vision, pp. 71–77. doi:10.1109/WACV.2015.17.

Milan, A., Leal-Taixe, L., Reid, I., Roth, S., Schindler, K., 2016. Mot16: A benchmark for multi-object tracking. URL: https://arxiv.org/abs/1603.00831, arXiv:1603.00831.

Neethirajan, S., 2022. Chicktrack – a quantitative tracking tool for measuring chicken activity. Measurement 191, 110819. URL: https://www.sciencedirect.com/science/article/pii/S0263224122001154, doi:https://doi.org/10.1016/j.measurement.2022.110819.

Novas, R.V., Usberti, F.L., 2017. Live monitoring in poultry houses: A broiler detection approach , 216–222URL: https://doi.org/10.1109/SIBGRAPI.2017.35, doi:10.1109/SIBGRAPI.2017.35.

Ojo, R.O., Ajayi, A.O., Owolabi, H., Oyedele, L.O., Àkànbí, L.A., 2022. Internet of things and machine learning techniques in poultry health and welfare management: A systematic literature review. Comput. Electron. Agric. 200, 107266. URL: https://doi.org/10.1016/j.compag.2022.107266, doi:10.1016/j.compag.2022.107266.

Okinda, C., Nyalala, I., Korohou, T., Okinda, C., Wang, J., Achieng, T., Wamalwa, P., Mang, T., Shen, M., 2020. A review on computer vision systems in monitoring of poultry: A welfare perspective. Artificial Intelligence in Agriculture 4. doi:10.1016/j.aiia.2020.09.002.

OpenCV, 2023. The opencv reference manual, Itseez. URL: https://docs.opencv.org/4.x/da/d13/tutorial_aruco_calibration.html.

Peña Fernández, A., Norton, T., Tullo, E., van Hertem, T., Youssef, A., Exadaktylos, V., Vranken, E., Guarino, M., Berckmans, D., 2018. Real-time monitoring of broiler flock's welfare status using camera-based technology. Biosystems Engineering 173, 103–114. doi:10.1016/j.biosystemseng.2018.05.008.

Peña Fernández, A., Norton, T., Tullo, E., van Hertem, T., Youssef, A., Exadaktylos, V., Vranken, E., Guarino, M., Berckmans, D., 2018. Real-time monitoring of broiler flock's welfare status using camera-based technology. Biosystems Engineering 173, 103–114. URL: https://www.sciencedirect.com/science/article/pii/S153751101730483X, doi:https://doi.org/10.1016/j.biosystemseng.2018.05.008. advances in the Engineering of Sensor-based Monitoring and Management Systems for Precision Livestock Farming.

du Plessis, E.W., Beausoleil, N.J., Bolwell, C.F., Stafford, K.J., 2021. Validation of a combined approach-avoidance and conditioned stimulus aversion paradigm for evaluating aversion in chickens. Plos one 16, e0247674.

Riber, A.B., Wurtz, K.E., 2024. Impact of growth rate on the welfare of broilers. Animals 14, 3330. doi:10.3390/ani14223330.

Ristani, E., Solera, F., Zou, R.S., Cucchiara, R., Tomasi, C., 2016. Performance measures and a data set for multi-target, multi-camera tracking. URL: https://arxiv.org/abs/1609.01775, arXiv:1609.01775.

Santos, J.L.d., Amoroso, L., Silvério, F.C., Costa, A.H.A., Silva, M.C., Artoni, S.M.B., Costa, G.C., Rubio, M.S., 2024. Musculoskeletal and visceral quality of broilers with different body patterns. Brazilian Journal of Poultry Science 26. doi:10.1590/1806-9061-2023-1880.

Tang, Z., Lin, Y.S., Lee, K.H., Hwang, J.N., Chuang, J.H., 2019. Esther: Joint camera self-calibration and automatic radial distortion correction

from tracking of walking humans. IEEE Access 7, 10754–10766. doi:10.1109/ACCESS.2019.2891224.

Teepe, T., Wolters, P., Gilg, J., Herzog, F., Rigoll, G., 2024. Lifting multi-view detection and tracking to the bird's eye view, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 667–676.

Trocino, A., White, P., Bordignon, F., Ferrante, V., Bertotto, D., Birolo, M., Pillan, G., Xiccato, G., 2020. Effect of feed restriction on the behaviour and welfare of broiler chickens. Animals 10. URL: https://www.mdpi.com/2076-2615/10/5/830, doi:10.3390/ani10050830.

Tzutalin, 2015. Labelimg https://github.com/tzutalin/labelimg. URL: https://github.com/tzutalin/labelImg.

Van der Eijk, J., Guzhva, O., Voss, A., Möller, M., Giersberg, M., Jacobs, L., De jong, I., 2022. Seeing is caring – automated assessment of resource use of broilers with computer vision techniques. Frontiers in Animal Science 3. doi:10.3389/fanim.2022.945534.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods 17, 261–272. doi:10.1038/s41592-019-0686-2.

Vora, J., Dutta, S., Jain, K., Karthik, S., Gandhi, V., 2023. Bringing generalization to deep multi-view pedestrian detection, in: IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, WACV 2023 - Workshops, Waikoloa, HI, USA, January 3-7, 2023, IEEE. pp. 110–119. URL: https://doi.org/10.1109/WACVW58289.2023.00016, doi:10.1109/WACVW58289.2023.00016.

Wang, C., Wang, Y., Wang, Y., Wu, C.T., Yu, G., 2019. mussp: Efficient min-cost flow algorithm for multi-object tracking, in: Advances in Neural Information Processing Systems, pp. 423–432.

Xu, Y., Li, Y.J., Weng, X., Kitani, K., 2021. Wide-baseline multi-camera calibration using person re-identification, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13129–13138. doi:10.1109/CVPR46437.2021.01293.

Yang, X., Chai, L., Bist, R.B., Subedi, S., Wu, Z., 2022. A deep learning model for detecting cage-free hens on the litter floor. Animals 12. URL: https://www.mdpi.com/2076-2615/12/15/1983, doi:10.3390/ani12151983.

Zaidi, S.S.A., Ansari, M.S., Aslam, A., Kanwal, N., Asghar, M.N., Lee, B., 2022. A survey of modern deep learning based object detection models. Digit. Signal Process. 126, 103514. URL: https://doi.org/10.1016/j.dsp.2022.103514, doi:10.1016/j.dsp.2022.103514.

Zhang, L., Gao, J., Xiao, Z., Fan, H., 2023. Animaltrack: A benchmark for multi-animal tracking in the wild. Int. J. Comput. Vis. 131, 496–513. URL: https://doi.org/10.1007/s11263-022-01711-8, doi:10.1007/S11263-022-01711-8.

Zhu, C., 2019. Multi-Camera People Detection and Tracking. Master's thesis. KTH, School of Electrical Engineering and Computer Science (EECS).