animals

MDPI

Article

Comparison of the Unrest Index with Tracker-Derived Travel Distance and Optical Flow in Broiler Videos

Igor Cristian Hübner [1], Allan Lincoln Rodrigues Siriani [2], Carlos Nobuaki Hokumura Filho [2], Mário Mollo Neto [2], Daniella Jorge de Moura [1,*] and Juliana de Souza Granja Barros [1]

[1] School of Agricultural Engineering (FEAGRI), University of Campinas (UNICAMP), Campinas 13083-875, SP, Brazil; igor.hubner2020@gmail.com (I.C.H.)
[2] Department of Management, Development and Technology, School of Sciences and Engineering, São Paulo State University (UNESP), Tupã 17602-496, SP, Brazil; allan.siriani@unesp.br (A.L.R.S.); mario.mollo@unesp.br (M.M.N.)
* Correspondence: djmoura@unicamp.br

Simple Summary

How closely do video measures agree on when a group of chickens moves most? We compared the Unrest Index, which measures how much bird positions change between images, with the distance walked by tracked birds and with motion across the image. In eight recordings of groups of ten broilers, the index agreed closely with tracked distance, but both calculations use the same position estimates, so part of that agreement is built in. Agreement with image motion, which does not depend on the bird detector, was moderate. When the same detected birds were used, locating each bird by the center of its outline instead of the center of its box made no consistent difference; the better agreement seen earlier came from using a different model. Using the 95th percentile of the changes instead of the largest change barely improved agreement. Measuring how far birds actually walk will require independently checked paths.

Abstract

We compared the stepwise Unrest Index with tracker-derived travel distance and dense optical flow in eight overhead recordings of groups of ten Ross broilers recorded 15–40 days after housing. The principal recording lasted 29.8 min and the other seven about 60 min; four supplied no frames to model training or validation. In each recording, one tracked pass of an instance-segmentation model gave raw box centers, mask-area centroids and tracker box centers for identical instances. Agreement was Spearman's ρ with paired block bootstrap intervals. With instances held fixed, the mask centroid showed no consistent difference from the box center in agreement with optical flow ($\Delta\rho$ −0.048 to +0.051). The higher agreement previously attributed to mask centroids was reproduced only when the model changed ($\Delta\rho$ +0.042 to +0.120; all within-recording intervals above zero). The 95th-percentile Hausdorff distance gave small gains over the maximum ($\Delta\rho$ +0.020 to +0.031 against flow), which do not identify an optimal percentile. Agreement with travel distance ($\rho = 0.77$–0.89) was stronger than with optical flow ($\rho = 0.28$–0.59), and switching the travel-distance coordinates from mask centroids to raw box centers reduced the correlation by 0.15–0.23. Restricting flow to the pen floor raised index–flow agreement. Distances are in pixels, and all results concern about ten birds per frame.

Keywords: broilers; Unrest Index; Hausdorff distance; optical flow; instance segmentation; multi-object tracking

Check for updates

Academic Editor: Sabine Gebhardt-Henrich

Received: 26 October 2026
Revised: 10 November 2026
Accepted: 16 November 2026
Published: 18 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Heat stress is a major constraint on poultry production, reducing performance and increasing mortality, with substantial economic losses [1,2]. Because each species expresses discomfort through characteristic changes in behavior, behavior can be used to infer an animal's welfare state [1,3]. In commercial houses containing thousands of birds at high stocking densities, individual welfare assessment is impracticable. This has motivated automated surveillance systems within precision livestock farming (PLF) [4,5].

Two computer vision approaches using ordinary camera images can extract flock movement metrics without requiring individual identification. Optical flow summarizes frame-to-frame brightness changes using displacement statistics such as mean, variance, skewness, and kurtosis [6,7]. The Unrest Index uses the Hausdorff distance between sets of bird centroids in consecutive frames to quantify how much the group has departed from its stationary configuration [1]. Both metrics depend on the frame-sampling rate: displacement observed between samples decreases as sampling frequency increases [8].

Del Valle et al. [1] originally proposed the Unrest Index using classical background-subtraction segmentation. Its practical use has since expanded. Pereira et al. [9] applied it to gait assessment in broilers, Fernandes et al. [10] used it to estimate laying hen welfare under different thermal and lighting conditions, and Pereira et al. [8] showed that it could be calculated from the bounding-box centers of a YOLO detector without segmentation. The latter study found strong equivalence with tracker-derived travel distance and optical-flow statistics, reporting Pearson correlations of 0.86–0.96, depending on strain, between values aggregated by experimental day.

At the individual time-step level, however, classical Hausdorff distance has a structural weakness that has received little attention in poultry research. It is a worst-case statistic defined by a supremum, or in practice a maximum, over points [11,12]. One missed bird, a spurious detection in the background, or a centroid displaced by partial occlusion in a single frame can dominate the entire index value for that step even if every other bird remained still. In a real video, motion blur, occlusion by equipment, and birds entering or leaving the scene produce a heavy-tailed series. Rank correlation is therefore used for the ranking of time steps, and product-moment correlation is reported for the linear association of magnitudes (Section 4.1). These properties also motivate re-examination of two design choices: the source of the centroid and the Hausdorff-distance variant.

The centroid is a second design choice when the index is computed from a detector. Pereira et al. [8] used the bounding-box center, whereas the original definition [1] used the centroid of the segmented bird area, which deep instance segmentation can now recover. The two positions differ for elongated or partly occluded birds, but whether either measures locomotion more accurately has not been tested.

Classical computer vision offers a way to make the worst-case statistic more robust. Partial, or percentile-based, Hausdorff distance replaces the maximum with the kth percentile of nearest-neighbor distances [11–13].

1.1. Background and Related Work

1.1.1. Flock Movement and Optical Flow

Behavior is among the most sensitive and least invasive indicators of animal welfare [14]. Locomotor activity, in particular, is used as an indicator in broilers, whose mobility is compromised by selection for rapid growth [3,8]. PLF operationalizes this premise using sensors and data analysis [4,15,16], with the caveat that measured variables must have biological rather than merely operational meaning [17]. A systematic review found that most poultry PLF solutions were designed to monitor production and the environment rather than welfare [18]. Among signals available from overhead cameras, which

are inexpensive and compatible with existing digital video recorder (DVR) systems, flock movement dynamics stand out because they do not require individual identification [5,19].

The best-established approach is optical flow, developed by M. S. Dawkins and colleagues. Higher-order statistics of the aggregated flow field (mean, variance, skewness, and kurtosis) correlate with gait score, footpad lesions, and mortality [20], discriminate flock behavior [6], have established behavioral correlates [7], model welfare outcomes through Bayesian regression [21], and anticipate Campylobacter outbreaks [22]. Gebhardt-Henrich et al. [23] linked flock flow patterns to the behavior of individual birds. Related approaches include activity-deviation detection in groups [24] and depth-camera quantification of inactivity [25], supported by dense-flow estimators such as two-frame polynomial expansion [26]. A structural limitation of optical flow is that it measures aggregate movement magnitude in the image plane: it is sensitive to lighting and shadows, depends on parameter settings, does not isolate individuals or yield trajectories, and is not invariant to density.

1.1.2. The Unrest Index and Bird Tracking

A second family of metrics developed by D. F. Pereira's group derives indicators from bird positions. The cluster index quantifies spatial aggregation as a proxy for thermal stress [27]. The Unrest Index was defined as the Hausdorff distance between centroid sets in consecutive frames: greater spatial rearrangement of the flock produces a higher index [1]. Del Valle et al. [1] obtained centroids through classical segmentation and showed that the index detects restlessness under thermal conditions. Fernandes et al. [10] applied it to laying hens under different thermal and lighting conditions, Massari et al. [28] combined vision-based indices as a proof of concept, and Pereira et al. [9] associated walkway movement with gait score. The closest study is Pereira et al. [8], which compared optical flow, the Unrest Index, and travel distance in three strains using a You Only Look Once (YOLO) detector, box centers as centroid approximations, and the maximum of the two directed Hausdorff distances. An equivalent formulation appears in SY-Track [29]. In all these studies without exception, the index uses the distance supremum.

Position estimation increasingly relies on YOLO detectors [30–32], trained on datasets annotated with tools such as the Computer Vision Annotation Tool (CVAT) [33]. In poultry applications, ChickTrack combined YOLO with a Kalman filter ([34]; reviewed by [35]). Siriani et al. [36] adapted YOLO to low light; Siriani et al. [37] coupled YOLO with BoT-SORT to classify activity states; and SY-Track used YOLO with a SORT-like tracker [29]. Detection-association trackers such as ByteTrack [38] are now standard. Doornweerd et al. [39] concluded that bird kinematics do not explain identity switches; the bottleneck is the near-identical appearance of individuals. Mehdizadeh et al. [40] mitigated this with segmented body area, and van der Eijk et al. [41] used Mask R-CNN with zone classifiers. A recurrent step in these studies and in work on the Unrest Index is manual delineation of the region of interest in the first frame [1,8]. Siriani et al. [42] automate this step with a dynamic region of interest that follows pendular equipment movement.

1.1.3. Partial Hausdorff Distance and the Research Gap

Because it is a supremum, undirected Hausdorff distance is determined by the worst-matched point pair. A single unmatched point caused by occlusion, segmentation noise, or a spurious detection can dominate the metric [11]. Recognizing this weakness, the original authors introduced partial Hausdorff distance, replacing the maximum by the kth-largest nearest-neighbor distance (a quantile) so that a fraction of unmatched points can be tolerated [11]. Dubuisson and Jain [12] found that modified Hausdorff distance, based on a mean rather than a supremum, is more robust under noise. In medical-image segmenta-

tion, the 95th percentile (HD95) has been established practice for more than a decade [13]. This replacement is standard in computer vision but has not been transferred to the poultry Unrest Index.

Previous studies used the maximum nearest-neighbor distance [1,8,10,28,29], although partial Hausdorff measures are available [11–13]. Existing poultry comparisons provide limited evidence on how centroid source and percentile choice affect the ranking of individual video time steps.

The objective was to quantify the stepwise rank agreement of the Unrest Index with tracker-derived travel distance and optical flow at approximately ten birds per frame to separate the effect of the centroid definition from that of the model by deriving both centroids from identical segmentation instances and to describe how agreement varies with the Hausdorff percentile.

2. Materials and Methods

2.1. Overview and Dataset

From an overhead video of a broiler compartment, each sampled frame is processed by YOLO26 models trained for the single class bird. They return a bounding box (or segmentation mask) and a tracking identifier for each bird. Three flock movement metrics are calculated on the same timeline: the Unrest Index (Hausdorff distance between centroids in consecutive frames), in its classical and percentile-based partial variants; tracker-derived flock travel distance summed over identifiers; and dense Farnebäck optical-flow statistics, which are independent of the detector. The series are compared by rank correlation (Section 2.7). Two pipelines were first compared: a detector read at the box center and a segmenter read at the mask-area centroid (Section 2.5). Because these pipelines differ in model as well as in centroid, a controlled comparison then derived every centroid definition from the same segmentation instances (Section 2.8).

The models were trained on frames randomly extracted from four recordings acquired under different conditions (resolution, lighting, camera and bird age); resolutions were 960 × 1080 and 1280 × 720 pixels. Boxes and masks were manually annotated and reviewed frame by frame in CVAT [33] for the single class bird. The segmenter used all 1528 annotated frames and the detector an earlier, fully reviewed subset of 773 frames. Frames were assigned at random to validation with a ratio of 0.15 (seed 42), giving 1299 training and 229 validation frames for the segmenter and 657 and 116 for the detector (Table 1). This frame-level split places temporally adjacent frames from the same recording in both sets and includes frames from the principal recording and three replication recordings in training; this leakage is discussed in Section 4.3. The other four recordings supplied no frames to either model.

Table 1. The eight recordings. Date and time are from the recorder clock at the start of the analyzed segment; time since housing is in days. Compartment (Comp.) is from the camera label in the image (CAM 1–3 = C1–C3 in all batches). Hz: processed frame rate. Seg.: segmenter training/validation frames; Det.: detector frames (training + validation). † No frames in either model's development pool. SimPlay-CAM2 shows the same birds and compartment as RecordPlan-02/24a; the three recordings of 18 June were acquired in the same session, and the two of 24 February on the same day.

Recording	Date, Time	Batch	Comp.	Treatment	Days Since Housing	Resolution (px)	Hz	Seg. Train/Val	Det.
DVR-CAM2 (principal)	18 June, 08:00	3	C2	Thermoneutral	36	960 × 1080	13.2	459/73	492
RecordPlan-04/23	23 April, 10:30	2	C3	Heat stress	29	1280 × 720	10.0	571/99	118

Table 1. *Cont.*

Recording	Date, Time	Batch	Comp.	Treatment	Days Since Housing	Resolution (px)	Hz	Seg. Train/Val	Det.
RecordPlan-02/24a	24 February, 10:01	1	C2	Heat stress	20	960 × 1080	10.0	163/28	148
RecordPlan-02/24b	24 February, 14:02	1	C1	Thermoneutral	20	1280 × 720	10.0	106/29	15
RecordPlan-04/09 †	9 April, 08:44	2	C1	Thermoneutral	15	1280 × 720	10.0	0	0
SimPlay-CAM2 †	16 March, 11:52	1	C2	Heat stress	40	960 × 1080	10.0	0	0
DVR-Ch2 †	18 June, 09:12	3	C1	Heat stress	36	1280 × 720	9.9	0	0
DVR-Ch3 †	18 June, 09:11	3	C3	Cold stress	36	1280 × 720	9.8	0	0
Total								1299/229	773

2.2. Experimental Facilities and Video Acquisition

The recordings come from an experiment at the Laboratory of Thermal Comfort II (LCT II), School of Agricultural Engineering, University of Campinas (FEAGRI/UNICAMP), Campinas, Brazil, using three separately managed compartments (C1–C3) within one climate chamber, each housing 10 Ross broilers. The litter floor measured 156 × 130 cm (2.03 m^2) in C1 and C2 and 156 × 134 cm (2.09 m^2) in C3, giving 0.203–0.209 m^2 per bird (4.8–4.9 birds/m^2). Three 42-day batches were reared from 4 February to 18 March, 25 March to 6 May, and 13 May to 24 June 2026, with 90 birds across the broader experiment. Within each batch, the thermoneutral, heat-stress and cold-stress treatments were randomly allocated to the compartments, giving a randomized complete block design with batch as block (batch 1: C1 thermoneutral, C2 heat, C3 cold; batch 2: C1 thermoneutral, C2 cold, C3 heat; batch 3: C1 heat, C2 thermoneutral, C3 cold). Treatment is reported for completeness (Table 1) and does not enter any analysis; this article evaluates movement metrics rather than treatment effects. Lighting and ventilation followed the Ross management manual, subject to the conditions imposed by each treatment. Lights were on throughout every recording, there was no entry or handling during the recordings, ten birds were alive in every compartment on every recording day, and no bird moved between compartments.

Video was recorded with one HiLook THC-T110-P analog Turbo HD turret camera(Hikvision, Hangzhou, China). (1 MP) in each of the three compartments, mounted at the center of the ceiling and connected to an Intelbras DVR(Intelbras S.A., São José, Brazil); Intelbras SIM Next stored the recordings on a dedicated 20 TB drive in 60 min files. Measured with a plumb line, the lens was 212 cm above the floor in all three compartments (about 202 cm above the 10 cm wood-shaving litter). Each camera covered the whole floor, including the feeder and the nipple drinker; a field check with nine floor markers found all of them visible in all three cameras. Each camera remained over the same compartment in all three batches. Acquisition date and time were verified against the clock printed by the recorder on the image and the compartment against the camera label in the image; the channel number in the file name does not identify the compartment. Time since housing at recording ranged from 15 to 40 days (Table 1). No behaviors were scored and no ethogram was used: the index uses all detected bird positions at each sampled step, so changes in posture or outline can contribute alongside locomotion.

2.3. Detection and Segmentation Models

Two single-class YOLO26m models (Ultralytics version 8.4.51; Ultralytics Inc., Frederick, MD, USA; [32]) were used: a detector (box; 100 epochs) and an instance segmenter (mask; 150 epochs). Both started from Common Objects in Context (COCO) pretrained weights, with batch size 8, imgsz = 640, automatic optimizer (lr0 = lrf = 0.01), cos_lr off, close_mosaic = 10, standard Ultralytics augmentation, seed 0, and deterministic execution; the segmenter additionally used overlap_mask and mask_ratio = 4. Table 2 gives the validation performance (segmenter: 229 images, 2278 instances; detector: 116 images). Complete hyperparameters, software versions, and hardware (NVIDIA RTX 5060 Ti GPU, 16 GB; CUDA 12.8; PyTorch 2.11) appear in Table 3. The method is architecture-agnostic: any detector/segmenter providing boxes and masks can be used.

Table 2. Single-class bird model performance on the validation set. These in-distribution metrics share source videos with training (Table 1) and do not measure generalization to new houses (Section 4.3). mAP, mean average precision at an IoU threshold of 0.5 (mAP@0.5) and averaged over thresholds 0.5–0.95 (mAP@0.5:0.95).

Model	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Detector (box)	0.984	0.984	0.994	0.947
Segmenter v3 (box)	0.992	0.990	0.995	0.965
Segmenter v3 (mask)	—	—	0.994	0.897

Table 3. Computing environment and main hyperparameters.

Component	Specification
Hardware/CUDA	NVIDIA RTX 5060 Ti 16 GB; CUDA 12.8; driver 590.48.01
Libraries	numpy 1.26.4; opencv-python 4.11.0; torch 2.11.0 + cu128; ultralytics 8.4.51; scipy 1.17
YOLO26m (both models)	COCO initial weights; batch 8; imgsz 640; optimizer = auto; lr0 = lrf = 0.01; momentum 0.937; weight_decay 5×10^{-4}; warmup_epochs 3; cos_lr off; close_mosaic 10; seed 0; deterministic execution
Detector	100 epochs
Segmenter	150 epochs; overlap_mask; mask_ratio 4
Tracking	model.track (default Ultralytics ByteTrack); conf $\geq$ 0.40 reapplied to output; track_high/low_thresh 0.25/0.10; new_track_thresh 0.25; match_thresh 0.80; track_buffer 30
Optical flow	calcOpticalFlowFarneback: pyr_scale 0.5; levels 3; winsize 15; iterations 3; poly_n 5; poly_sigma 1.2; frames at 0.5×; 8 × 8-pixel blocks
Partial Hausdorff	numpy.percentile (linear interpolation between order statistics)

The Segment Anything Model (SAM) 3.1 [43] served only to seed the masks. The full segmentation dataset, like the box dataset, was reviewed and corrected frame by frame in CVAT [33], so that both models' training labels are human-verified annotations rather than automatic output, and Table 2 measures agreement with human annotations. Earlier segmenter iterations were seeded by prompting SAM 3.1 within each bird's box, based on an untested assumption that a full-frame text prompt would have a recall ceiling. Rerunning it on full frames recovered a mean 9.94 masks per frame (versus 3.03 in the

previous set). New instances were merged with manual corrections (intersection over union, IoU, <0.3 indicating a new instance), and seg v3 was trained on the corrected set (15,185 instances).

2.4. Sampling, Region of Interest, and Detection

Del Valle et al. [1] and Pereira et al. [8] manually crop the area of interest in each scene (walls, pipes, windows, neighboring pens and highly reflective regions) before calculating metrics. The primary analysis here processed the full frame. This admits possible detections at pen boundaries into F(i) and includes the area outside the pen and the recorded timestamp in the optical-flow field. As a sensitivity analysis, a fixed polygon covering the pen floor was therefore defined for each camera (Section 2.8). The full frame was kept as the primary analysis because it was the specified analysis and the polygons were drawn after the primary results were known.

Each selected frame was processed with model.track() using the built-in Ultralytics ByteTrack configuration [38] (Table 3). The confidence threshold conf = 0.40 was applied at non-maximum suppression, before the tracker. ByteTrack associates detections with tracks by the overlap between detection boxes and Kalman-predicted track boxes, first for high-confidence detections and then, in a recovery stage, for detections scoring between track_low_thresh (0.10) and track_high_thresh (0.25). Because every retained detection scored at least 0.40, no detection reached the recovery stage. Masks and mask centroids do not enter association; they are computed after it. The boxes returned by the tracker are taken from its Kalman-filter state rather than from the raw detections, whereas each mask belongs to the raw detection. In the two pipelines, the box center was the center of this tracker box; the controlled comparison (Section 2.8) records both.

One in every three native frames was processed. Del Valle et al. [1] used 1 Hz, arguing that higher frequencies register negligible distances, a claim specific to a metric without identities. Here, tracker-derived travel distance (Section 2.6) depends on identity continuity between processed frames: when sample-to-sample displacement is large, overlap-based (IoU) association often fails, fragmenting identities and underestimating distance. All three metrics use this same timeline without resampling. Sensitivity to sampling rate was not characterized and is future work.

2.5. Unrest Index: Centroid and Partial Hausdorff Distance

In the original definition ([1], Equation (3)), the centroid of a bird is the mean position of the pixels in its connected blob B,

$$(\bar{x},\ \bar{y}) = \left(|B|^{-1}\ \Sigma[p \in B]\ x_{\mathrm{p}},\ |B|^{-1}\ \Sigma[p \in B]\ y_{\mathrm{p}}\right). \quad (1)$$

The blob was obtained by background subtraction against a static, clean background. We compare two centroid sources, each entering Equation (4) onward identically. The first, the box center (the approximation in [8]), is c = $((x_1 + x_2)/2, (y_1 + y_2)/2)$. It is simple and detector-independent but biased for elongated or asymmetric poses, without an occlusion-robustness advantage (the box simply expands and moves its center away from the visible body mass). The second, the mask-area centroid, uses seg v3 and the first-order moment of the segmented region, $((\bar{x},\ \bar{y})) = (\mathrm{m}_{10}/\mathrm{m}_{00}, \mathrm{m}_{01}/\mathrm{m}_{00})$, calculated with cv2.moments on the predicted polygon. It is the exact mean position of the occupied area, faithful to Equation (1) for the visible predicted region. Its error relative to actual movement remains unmeasured.

Let $F(i)$ and $F(i-1)$ denote centroid sets in consecutive sampled frames. The directed and symmetric Hausdorff distances are [1,11]

$$h_A(B) = sup[p \in A]\ inf[q \in B]\ \|p - q\|, \quad (2)$$

$$d_H(A, B) = max(h_A(B),\ h_B(A)), \tag{3}$$

and the Unrest Index at step i is

$$II(i) = d_H(F(i),\ F(i-1)). \tag{4}$$

All distances are expressed in image pixels. No image-to-floor calibration was performed, so distances are not converted to centimeters and are not comparable between recordings of different resolution.

Classical Hausdorff distance is a worst-case statistic: one missed or spurious detection can dominate II(i). Replacing the maximum with a percentile or a measure of central tendency is well established in computer vision, from partial Hausdorff distance [11] and modified Hausdorff distance [12] to routine use of the 95th percentile (HD95) in medical images [13]. We use the partial form, in which each directed distance is the kth percentile of nearest-neighbor distances, with linear interpolation between order statistics:

$$d_H(A, B;\ k)\ =\ max(P_k\{inf[q \in B]\ \|p\ -\ q\|\ :\ p \in A\},\ P_k\{inf[p \in A]\ \|p\ -\ q\|\ :\ q \in B\}). \tag{5}$$

The classical form is the maximum rather than a specified percentile. The classical, $k = 95$, and $k = 50$ (median) variants are reported in detail, and a full sweep $k \in \{30, ..., 95\}$ (Section 3.3) is performed for both centroid sources and repeated across eight videos (Section 3.4).

2.6. Comparison Metrics

Tracker-derived travel distance. For each identifier, centroid displacement is summed from frame to frame over every step in which that bird was observed (Section 2.2 of [8]):

$$\mathrm{d_{ID}} = \Sigma[i = \mathrm{f_2},...,\mathrm{f_n}]\ \surd((x_i - x_{i-1})^2 + (y_i - y_{i-1})^2). \tag{6}$$

Here, $\mathrm{f_2}$ and $\mathrm{f_n}$ are the second and last steps at which the identifier was observed. Identity turnover is tolerated as in the originating study: a lost identifier ceases accumulation and a new one begins from zero. As an implementation improvement, a gap in an identifier's presence between consecutive processed steps initiates a new accumulation segment instead of summing a straight-line displacement ("teleportation"). The comparison uses flock distance per step (the increment), not the cumulative sum, which would yield spurious correlation through a shared monotonic trend. Because this metric and the Unrest Index use the same detections and centroids, they are not fully independent (Sections 2.8 and 3.5).

Dense Farnebäck optical flow [26] was calculated between successive processed frames after downscaling by a factor of 0.5 (Table 3). We computed the mean flow magnitude in each 8 × 8-pixel block and summarized the distribution of block means by its mean, variance, skewness and kurtosis [7,20]. The mean was used as the comparator. It was selected after inspecting agreement on these same recordings, so agreement with it is exploratory rather than confirmatory; correlations with all four statistics are reported for every recording (Table S1). Full-frame flow also responds to lighting, equipment, motion outside the pen and the timestamp [8]; Section 2.8 describes a region-of-interest sensitivity analysis.

2.7. Within-Recording Comparisons and Recording-Level Replication

For each series in Table 4, the median and interquartile range (IQR) summarize valid processed steps in the principal recording. The IQR is the 75th percentile minus the 25th percentile, in the units of the metric. Observations were paired by the processed time step within each recording. Spearman's ρ describes agreement in the ranking of time steps, and Pearson's r describes linear association in the original units. Confidence intervals were obtained with a moving-block bootstrap [44], which resamples contiguous blocks to

preserve temporal dependence. Blocks of b = round(L × f) steps, where f is the effective processed frame rate, were drawn uniformly from all n − b + 1 start positions without wrapping, concatenated and truncated to the series length n. Coefficients were recomputed on the re-ranked resample, with 2000 replicates, a fixed seed (20261104) and 95% intervals. For contrasts, the same resampled indices were applied to every series, so differences are paired. In the principal recording, estimated integrated autocorrelation times were about 7, 90 and 187 processed steps for the index, travel distance and flow (0.5, 7 and 14 s). The primary block length was therefore L = 90 s, which spans all three timescales; 10 s, 300 s and 600 s were sensitivity analyses. The main tables report intervals with the primary 90 s blocks; intervals with 10 s blocks for Table 5 and with 10, 300 and 600 s blocks for the controlled contrasts are in Tables S3 and S4. All intervals describe uncertainty within a recording. The eight recordings include same-day and same-flock sets (Table 1) and are not a sample of independent flocks or batches, so between-recording variation is summarized descriptively by the mean, the range and the number of recordings whose interval excludes zero.

Table 4. Descriptive statistics for the principal video (23,148 steps, ≈13 Hz). Unrest Index in pixels; IQR, interquartile range.

Metric	Median	IQR	p95	Maximum
Birds detected per frame	10	0	11	13
Optical flow (mean magnitude)	0.06	0.04	0.14	1.44
Travel distance per step [px]	3.84	4.54	15.7	138.7
Unrest Index—box, classical [px]	1.70	3.76	81.3	315.8
Unrest Index—box, p95 [px]	1.32	2.56	49.2	263.6
Unrest Index—mask, classical [px]	1.67	2.71	12.2	312.0
Unrest Index—mask, p50 [px]	0.19	0.15	0.61	15.4

Table 5. Stepwise correlations among flock movement metrics by centroid source and Hausdorff variant: Spearman (ρ) and Pearson (r), with 95% confidence intervals (CI) from a moving-block bootstrap with the primary 90 s blocks (2000 replicates, seed 0; Section 2.7). Intervals with 10 s blocks are given in Table S3.

Configuration	Metric Pair	ρ (95% CI)	r (95% CI)
Box + classical	II × optical flow	0.498 (0.443, 0.555)	0.108 (0.078, 0.144)
Box + classical	II × travel distance	0.798 (0.770, 0.831)	0.102 (0.077, 0.131)
Box + classical	optical flow × travel distance	0.629 (0.573, 0.683)	0.744 (0.694, 0.783)
Box + p95	II × optical flow	0.520 (0.466, 0.576)	0.116 (0.090, 0.152)
Box + p95	II × travel distance	0.822 (0.793, 0.855)	0.114 (0.091, 0.141)
Box + p95	optical flow × travel distance	0.629 (0.574, 0.684)	0.744 (0.694, 0.782)
Mask + classical	II × optical flow	0.576 (0.543, 0.615)	0.180 (0.147, 0.223)
Mask + classical	II × travel distance	0.892 (0.876, 0.905)	0.215 (0.180, 0.261)
Mask + classical	optical flow × travel distance	0.677 (0.644, 0.716)	0.793 (0.748, 0.834)
Mask + p50	II × optical flow	0.498 (0.443, 0.559)	0.683 (0.618, 0.737)
Mask + p50	II × travel distance	0.620 (0.584, 0.656)	0.710 (0.658, 0.763)
Mask + p50	optical flow × travel distance	0.677 (0.644, 0.716)	0.793 (0.748, 0.834)

Series co-monotonicity (Spearman relative to box + classical): box + p95, ρ = 0.992; mask + classical, ρ = 0.606; mask + p50, ρ = 0.343.

The principal recording is a 960 × 1080-pixel overhead DVR video of one compartment. Its stream duration was 1790 s (ffprobe); processing every third frame yielded 23,148 steps at ≈13 effective Hz (≈29.8 min). Replication (Section 3.4) adds seven recordings, three represented in training and four absent from it, each ≈60 min at 9.8–10.0 effective Hz and ≈36,000 steps. In the principal recording, 241 identifiers in the detector pipeline and 89 in the segmenter pipeline contributed at least one displacement (Section 3.2), with a mean of 10.0 detected birds per frame (range 7–13). Steps with F(i) or F(i − 1) empty were flagged and excluded (one step in the principal recording). The computing environment and main hyperparameters are in Table 3.

2.8. Controlled Centroid Comparison and Region of Interest

The two pipelines differ in model weights, training data, retained detections and identities as well as in centroid definition. To separate these factors, the segmentation model was run once more on each of the eight recordings with the same tracker configuration, and three positions were recorded for every tracked instance: the center of the raw detector box (raw), the area centroid of the same detection's mask (mask), and the center of the box returned by the tracker (tracker box), which corresponds to the box center of the detector pipeline. Frames, retained instances, the confidence threshold and track identifiers are therefore identical across the three read-outs. This pass reproduced the mask series of the segmenter pipeline (principal recording: $\rho = 0.576$ with optical flow and 0.892 with travel distance). The Unrest Index (classical and all percentiles) and travel distance were recomputed from each read-out.

Contrasts in Spearman's ρ were estimated with the paired bootstrap: the centroid effect (mask − raw, same model); the smoothing effect (tracker box − raw); the model effect (segmenter − detector, both read at the tracker box center), judged against optical flow, the only reference common to both runs; and the percentile effect (p95 − classical; p95 corresponds to the HD95 convention [13]). The centroid and percentile effects were evaluated against optical flow and against each read-out's own travel distance. To assess the sensitivity of index–distance agreement to the coordinate read-out, each index was also correlated with travel distance computed from a different read-out of the same instances, $\rho(\text{UI_a}, \text{WD_a}) - \rho(\text{UI_a}, \text{WD_b})$. This contrast measures sensitivity to the choice of coordinates used for travel distance. It does not isolate a causal contribution from shared errors, because the read-outs locate different points and retain shared detections and identifiers. A primary family of six contrasts per recording (the centroid effect under the classical and p95 distances and the p95 − classical effect for the mask read-out, each against optical flow and against travel distance) was adjusted with the Holm procedure [45] within each recording; the full percentile sweep is descriptive.

Regarding region-of-interest sensitivity, for each camera, a fixed polygon covering the pen floor and excluding the timestamp overlay and the area outside the pen was drawn on the median frame of the recording; the coordinates are in the data repository. Instances were kept when the mask centroid lay inside the polygon or within 15 px of its edge, and the same instances were used for every read-out. Optical flow within the polygon was computed from a coarser grid saved in this pass (mean flow magnitude in 16 × 16-pixel blocks of the half-resolution frame) as the mean of the blocks whose center lay inside the polygon; the mean over all blocks reproduced the full-frame flow mean ($\rho = 1.000$ in every recording). The paired comparisons were repeated over the same steps with 90 as blocks (seed 20261105).

3. Results

3.1. Model Performance and Descriptive Statistics

The YOLO26m detector reached a precision and recall of 0.984, an mAP@0.5 of 0.994, and an mAP@0.5:0.95 of 0.947 on the validation set (Table 2). The seg v3 instance segmenter used for mask centroids reached a mask mAP@0.5 of 0.994 and an mAP@0.5:0.95 of 0.897 (229 validation images, 2278 instances). Both are in-distribution metrics: training and validation share source videos, and the video analyzed below also contributed frames to both models' training (Section 4.3). They quantify performance against annotations from the same source-video pool. Over the ≈29.8 min video, the bird count per frame remained stable (mean 10.0, range 7–13) for both centroid sources. Figure 1 illustrates segmentation: the birds are mostly dispersed, and masks remain separate even where birds touch (upper left in the first frame). Degradation is concentrated in birds partly behind equipment or at the edge of the field (truncated mask in the third frame).



Figure 1. Examples of instance segmentation (seg v3) in three frames from the analyzed video: filled mask, outline, and area centroid (white point) for each bird, with instance count in the corner.

Table 4 summarizes the principal video's 23,148 processed steps (23,147 with a defined Unrest Index; the first has no previous frame) for the four compared configurations. Travel distance per step and optical flow do not depend on the Hausdorff variant. Between pipelines, flock travel distance changes little, but the number of identifiers contributing a displacement falls from 241 (detector) to 89 (segmenter); these totals reflect the two models (Section 3.2). The Unrest Index itself differs by orders of magnitude between configurations: its median falls from 1.70 px (box, classical) to 0.19 px (mask, p50), because the median nearest-neighbor distance among ≈10 points is much smaller than the maximum. The box index series is nearly identical for classical and p95 Hausdorff ($\rho = 0.99$; Table 5), whereas mask + p50 gives a qualitatively different series ($\rho = 0.34$ relative to box + classical).

3.2. Correlations and Numbers of Track Identifiers

Table 5 reports Spearman and Pearson correlations with moving-block bootstrap intervals using the primary 90 s blocks (10 s blocks: Table S3). Classical index Pearson correlations against both comparators were low (0.10–0.22). Extreme steps can strongly influence these estimates. The intervals describe uncertainty within the recording.

With classical Hausdorff distance, the index–distance correlation was 0.80 for the detector pipeline and 0.89 for the segmenter pipeline. The corresponding flow–distance correlations were 0.63 and 0.68. Both comparisons change the source of the detections as well as the centroid representation.

Lowering the percentile to the median reduces ranking agreement. Switching from classical to p50 Hausdorff with mask centroids lowers ρ(Index × travel distance) from

0.89 to 0.62 because the median cutoff discards some fine ordering. The role of percentile is examined in Section 3.3.

The detector and segmenter pipelines produced 241 and 89 identifiers with at least one displacement in the principal recording; in the single-model pass (Section 3.5), the tracker emitted 153 identifiers, 89 of which contributed a displacement. These are distinct tracker labels accumulated over the recording, not bird identities: with ten birds, additional labels may reflect track loss and reinitialization, false detections or duplicate detections. Because masks do not enter association (Section 2.4), the difference between 241 and 89 reflects the two detectors, and without independently labelled trajectories neither identity-switch rates nor errors in estimated travel distance can be inferred from the totals. The Unrest Index compares unordered sets of positions, so relabeling the same positions leaves it unchanged; missed detections can still change the sets. In the travel-distance calculation, only consecutive observations of the same identifier contribute a displacement (Section 2.6).

3.3. Percentile Sweep in Groups of Approximately Ten Birds

For ten sorted nearest-neighbor distances, linear interpolation places p95 at rank $1 + 9 \times 0.95 = 9.55$. Thus, both the ninth and tenth distances enter p95, and their spacing determines how far p95 lies below the maximum. This property of the calculation does not require an assumption about how accurately either statistic measures locomotion.

Table 6 and Figure 2 show the exploratory percentile sweep. For the principal recording, the box and mask pipelines gave correlations of 0.63 and 0.62 at p50, 0.82 and 0.92 at p95, and 0.80 and 0.89 at the maximum. The sweep is descriptive; only the p95 − classical contrast is tested with paired intervals and Holm adjustment (Section 3.5).

Table 6. Percentile k sweep: stepwise Spearman rank correlation (ρ) between the Unrest Index and travel distance per step for both centroid sources; "max" is classical Hausdorff distance.

Centroid Source	p30	p40	p50	p60	p70	p80	p90	p95	Max
Box (detector)	0.51	0.57	0.63	0.69	0.75	0.81	0.82	0.82	0.80
Mask (segmenter)	0.49	0.56	0.62	0.69	0.76	0.83	0.91	0.92	0.89

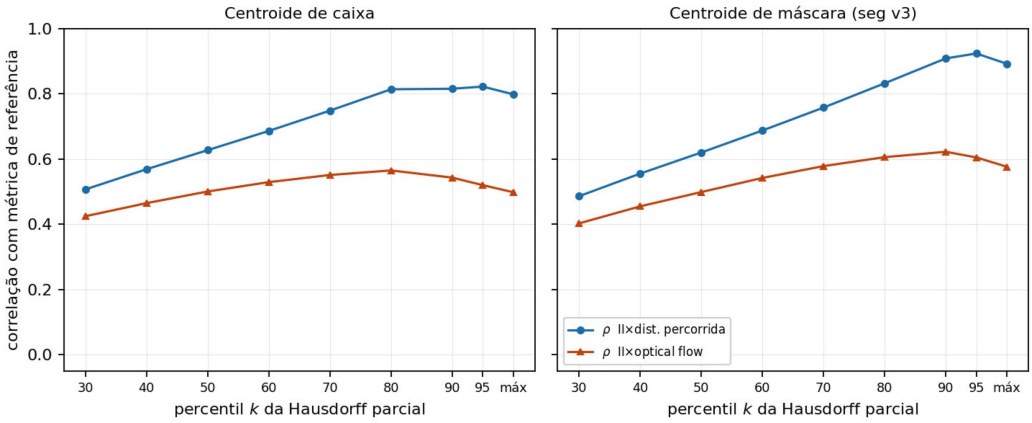


Figure 2. Spearman rank correlation (ρ) between the Unrest Index and travel distance per step (blue) and optical flow (orange) as a function of partial Hausdorff percentile k for both centroid sources; "max" indicates classical Hausdorff distance.

At p95, the index–distance correlation exceeded that for the classical maximum by 0.03 for masks and 0.02 for boxes (differences from rounded Table 6 values). Paired intervals for this contrast in all eight recordings are reported in Section 3.5.

For the classical index–distance correlation in the principal recording, the difference between the segmenter and detector pipelines was 0.09 (95% CI 0.06–0.12; 90 s blocks). This contrast changes the model as well as the centroid; Section 3.5 separates the two.

3.4. Replication Across Eight Videos

The full pipeline (percentile sweep, both pipelines) was repeated in seven further recordings, each ≈60 min and ≈36,000 steps, at 960 × 1080 or 1280 × 720 pixels (Table 1). Three contributed training frames (135–670 each) and four none. The held-out recordings differ in how independent they are: RecordPlan-04/09 shares neither birds nor a session with any training recording; SimPlay-CAM2 shows the same birds and compartment as RecordPlan-02/24a, which contributed training frames, 20 days later; and DVR-Ch2 and DVR-Ch3 show other compartments with different birds, recorded in the same session as the principal recording. Table 7 and Figure 3 summarize the results.

Table 7. Replication in eight broiler recordings: † denotes no frames in the models' development pool. Stepwise Spearman correlation (ρ) of the classical-Hausdorff Unrest Index with travel distance and optical flow for both pipelines (box: detector; mask: segmenter). Flow–walk denotes the association between the two comparators. Acquisition details are in Table 1; the median number of detected birds per frame was 10 in every recording.

Video	Centroid	Flow–Walk ρ	II × Distance ρ	II × Flow ρ
DVR-CAM2 (principal)	Box	0.63	0.80	0.50
DVR-CAM2 (principal)	Mask	0.68	0.89	0.58
RecordPlan-04/23	Box	0.67	0.60	0.46
RecordPlan-04/23	Mask	0.69	0.81	0.56
RecordPlan-02/24a	Box	0.32	0.82	0.25
RecordPlan-02/24a	Mask	0.35	0.85	0.28
RecordPlan-02/24b	Box	0.57	0.72	0.41
RecordPlan-02/24b	Mask	0.56	0.78	0.43
RecordPlan-04/09 †	Box	0.43	0.55	0.33
RecordPlan-04/09 †	Mask	0.47	0.77	0.36
SimPlay-CAM2 †	Box	0.44	0.76	0.34
SimPlay-CAM2 †	Mask	0.54	0.83	0.47
DVR-Ch2 (other pen) †	Box	0.64	0.63	0.48
DVR-Ch2 (other pen) †	Mask	0.67	0.87	0.56
DVR-Ch3 (other pen) †	Box	0.71	0.70	0.50
DVR-Ch3 (other pen) †	Mask	0.70	0.87	0.59

Rank agreement of the Unrest Index (mask centroids, classical Hausdorff) with travel distance replicated consistently: ρ = 0.77–0.89 step by step in all eight videos, including the four without training frames (ρ = 0.77, 0.83, 0.87, 0.87). The observed ranges overlapped; the four with training frames gave ρ = 0.78–0.89.

For the seven additional recordings in Table 7, the differences between pipelines in classical index–distance ρ were +0.21, +0.03, +0.06, +0.22, +0.07, +0.24 and +0.17 (from rounded coefficients). These differences combine model and centroid; Section 3.5 separates them with paired intervals. Agreement persisted in recordings that supplied no training frames, which excludes direct frame overlap as the explanation; shared birds, compartments or sessions (Table 1) may nevertheless contribute to the consistency.

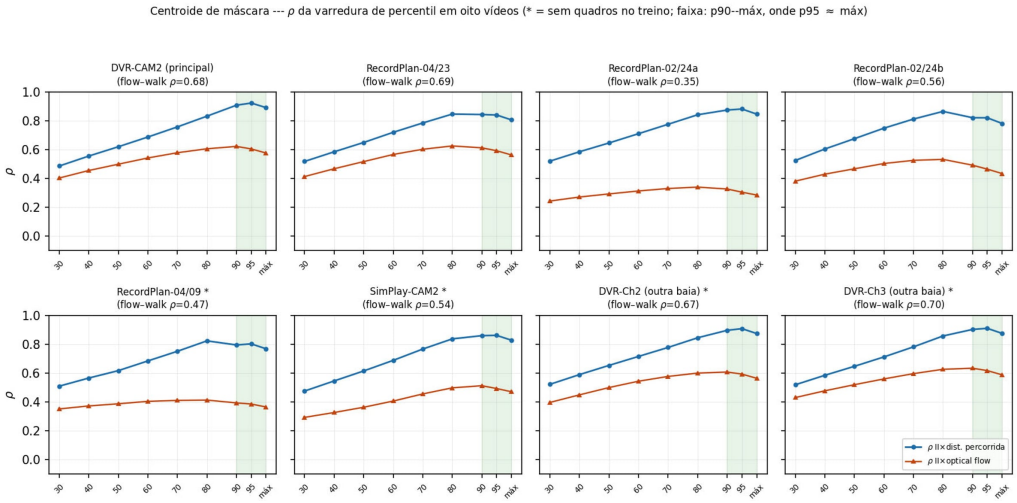


Figure 3. Spearman's rank correlations (ρ) between the Unrest Index calculated using mask centroids and tracker-derived travel distance (blue) or optical flow (orange), as a function of percentile k across eight videos. Asterisks identify videos that contributed no frames to model training. Correlations plateau between the 90th percentile (p90) and the maximum (max), with similar values at p95 and max.

Across all eight videos, the classical mask-based index correlated with mean optical flow at ρ = 0.28–0.59 and with tracker-derived travel distance at ρ = 0.77–0.89. Only flow is independent of the detections. Neither comparator provides independently observed bird displacement.

3.5. Controlled Comparison with Identical Instances

Table 8 summarizes the single-model pass. The median distance between the tracker and raw box centers was 0.08–0.14 px (95th percentile ≤ 0.34 px) and between the raw box center and the mask centroid 3.0–5.4 px. Against optical flow, the centroid effect (mask − raw) ranged from −0.048 to +0.051 across recordings (mean 0.000); its interval was above zero in two recordings, below zero in two and included zero in four. Against each read-out's own travel distance, the mask centroid gave slightly lower agreement (mean −0.011, range −0.018 to 0.000; interval below zero in seven recordings). Smoothing by the tracker had little effect against optical flow (mean +0.007, range −0.005 to +0.027). The model effect was positive in all eight recordings (mean +0.079, range +0.042 to +0.120; all intervals above zero) and of similar size to the contrast between the two original pipelines (mean +0.071 against flow). The higher agreement of the segmenter pipeline therefore reflects the change of model run (weights, training data and the resulting detections and identities), not the centroid definition.

p95 exceeded the classical distance in every recording and for every read-out (mask centroid against optical flow: mean +0.026, range +0.020 to +0.031; against travel distance: mean +0.035, range +0.032 to +0.039; all nominal intervals above zero, with significance reported under the Holm procedure). Replacing an index's own travel distance with travel distance from a different read-out lowered ρ by 0.19 on average for the mask and raw read-outs (range 0.15–0.23) and by 0.04–0.05 between the tracker and raw box centers, whose coordinates differ by a fraction of a pixel. With 10, 300 and 600 s blocks (Table S4) the point estimates were identical; the model effect, the p95 gain and the coordinate-read-out contrast remained above zero in all eight recordings, and the centroid effect again showed no consistent direction.

Table 8. Controlled comparison in identical segmentation instances (classical Hausdorff distance). Differences in Spearman ρ with paired 95% intervals (moving-block bootstrap, 90 s blocks, 2000 replicates). Centroid: mask − raw box center, against optical flow and against own travel distance; Model: segmenter − detector, both at the tracker box center, against optical flow; Shared: ρ(UI_mask, WD_mask) − ρ(UI_mask, WD_raw); p95: p95 − classical, mask read-out, against optical flow. * Significant under the reported Holm procedure for the primary family. The displayed intervals are nominal 95% bootstrap intervals. † No training frames.

Recording	Centroid vs. Flow	Centroid vs. Distance	Model vs. Flow	Shared	P95 vs. Flow
DVR-CAM2 (principal)	+0.020 (−0.002, 0.044)	−0.010 (−0.014, −0.005) *	+0.063 (0.031, 0.101)	+0.176 (0.152, 0.204)	+0.029 (0.025, 0.032) *
RecordPlan-04/23	−0.005 (−0.028, 0.020)	−0.017 (−0.026, −0.008) *	+0.117 (0.085, 0.152)	+0.181 (0.160, 0.204)	+0.029 (0.026, 0.033) *
RecordPlan-02/24a	−0.013 (−0.027, 0.005)	−0.010 (−0.014, −0.005) *	+0.042 (0.021, 0.065)	+0.148 (0.137, 0.161)	+0.020 (0.015, 0.025) *
RecordPlan-02/24b	−0.048 (−0.066, −0.031) *	−0.018 (−0.025, −0.010) *	+0.071 (0.031, 0.112)	+0.179 (0.155, 0.211)	+0.031 (0.027, 0.035) *
RecordPlan-04/09 †	−0.021 (−0.037, −0.005) *	−0.012 (−0.017, −0.007) *	+0.065 (0.033, 0.094)	+0.231 (0.210, 0.257)	+0.020 (0.017, 0.024) *
SimPlay-CAM2 †	+0.051 (0.019, 0.080) *	+0.000 (−0.009, 0.008)	+0.078 (0.039, 0.141)	+0.198 (0.172, 0.225)	+0.023 (0.018, 0.027) *
DVR-Ch2 †	+0.033 (0.008, 0.053) *	−0.007 (−0.014, −0.003) *	+0.075 (0.030, 0.113)	+0.212 (0.187, 0.240)	+0.029 (0.026, 0.033) *
DVR-Ch3 †	−0.020 (−0.043, 0.002)	−0.012 (−0.015, −0.007) *	+0.120 (0.077, 0.163)	+0.225 (0.190, 0.257)	+0.030 (0.025, 0.033) *
Mean (range)	0.000 (−0.048 to 0.051)	−0.011 (−0.018 to 0.000)	+0.079 (0.042 to 0.120)	+0.194 (0.148 to 0.231)	+0.026 (0.020 to 0.031)

3.6. Region-of-Interest Sensitivity

The pen-floor polygon removed one instance in one recording (SimPlay-CAM2) and none in the other seven, so the index and travel distance were practically unchanged. Restricting optical flow to the pen floor raised index–flow agreement in all eight recordings and for every read-out (Table S2). For the classical index with the mask centroid, ρ rose from 0.28–0.59 to 0.48–0.70 (Δρ +0.021 to +0.339; all intervals above zero), and p95 behaved the same way. The largest change occurred in RecordPlan-02/24a (0.28 to 0.62), where full-frame and pen-floor flow correlated at only ρ = 0.60, indicating substantial motion outside the pen in that view. Within the region of interest, the centroid effect against flow again showed no consistent direction (−0.065 to +0.073; interval above zero in one recording and below zero in three). Against pen-floor flow, the model effect remained positive in all eight recordings (mean +0.090, range +0.039 to +0.132; all intervals above zero), compared with +0.079 against full-frame flow.

4. Discussion

4.1. The Two Modifications

Spearman correlation addresses the intended ranking task. Pearson correlation addresses linear association of the measured magnitudes. The low classical index Pearson correlations in Table 5 limit the use of one series as a linear proxy for the other. A high rank correlation can persist despite occasional large false peaks, which would matter in an alarm application.

The segmentation pipeline gave higher index–distance and index–flow rank correlations than the detector pipeline in all eight recordings. The controlled comparison shows that this difference came from the model run rather than from the centroid definition: with identical instances and identifiers, the mask centroid and the raw box center gave agreement with optical flow that differed in sign between recordings and averaged zero, whereas changing only the model reproduced the original difference. The mask-area centroid implements the original area-based definition [1], but in these recordings it did not change agreement consistently. Which centroid better measures locomotion cannot be decided without annotated trajectories (Section 4.3).

At approximately ten birds per frame, upper percentiles preserved more agreement with tracked distance than the median. p95 gave a small but consistent gain over the maximum in all eight recordings and for every read-out ($\Delta\rho$ about +0.02 to +0.04). Whether a gain of this size matters in practice depends on the application, which this study does not assess.

Percentile-based distances are established in computer vision [11–13], but a percentile is a position in a finite set, so its effect depends on the number of points. With linear interpolation and n points, percentiles above approximately $k^*(n) \approx 100(1 - 1/n)$ depend only on the two largest nearest-neighbor distances (about 89–90 for $n = 10$), which explains why p95 behaves like the maximum at $n \approx 10$ (Section 3.3). Our recordings cover only that case. The relation neither identifies an optimal percentile nor predicts performance in larger groups, which would require recordings of groups of substantially different sizes.

At the observed group size, the number of detections helps interpret which nearest-neighbor distances contribute to a given percentile. The present recordings support reporting that count alongside the percentile, centroid definition and sampling interval. They do not establish a deployment threshold or an optimal percentile for other group sizes.

4.2. Alternative Interpretations and Comparison with the Literature

Three mechanisms can affect agreement. First, missed or spurious detections change the point sets and can affect both index and distance. Second, the index measures changes in a set of positions without preserving identity, whereas travel distance accumulates individual displacements. Third, lighting changes, equipment and the timestamp can contribute to full-frame optical flow [8]. The region-of-interest analysis supports this third mechanism: restricting flow to the pen floor raised index–flow agreement in every recording, by up to 0.34, without changing the comparison between centroid definitions. Agreement with full-frame flow is therefore lower than agreement with pen-floor optical flow; neither provides independent ground truth for bird movement.

Del Valle et al. [1] and Pereira et al. [8] reported stronger agreement under different conditions. Del Valle et al. [1] used a static, clean background, background-subtraction segmentation, 1 Hz sampling, and contrasts in thermal condition assessed by averages over minutes. Pereira et al. [8] used controlled camera and region of interest, a detector tuned to evaluated videos, and Pearson correlation between daily experimental aggregates rather than frame-to-frame increments. Our comparison is stepwise and rank-based, uses models trained for this recording type, and lacks camera calibration. It is a distinct, more demanding test. Here, the index (classical Hausdorff) in the segmenter pipeline agreed with travel

distance at $\rho = 0.77$–0.89 in all recordings, including those absent from training (Section 3.4), but part of this agreement comes from shared coordinates (Section 3.5). The magnitude equivalence reported for aggregates with product–moment correlation cannot be directly compared and was not sought here.

4.3. Limitations

The main limitations of this study are as follows. (i) No independently annotated trajectories were available, so neither comparator is ground truth for bird displacement, and agreement with travel distance, which shares detections and identifiers with the index, is not independent validation. (ii) All recordings come from one laboratory, with about ten birds per compartment recorded 15–40 days after housing; the results do not extend to larger groups, other houses, stocking densities or camera setups. (iii) Model metrics are in-distribution because of the frame-level split and the overlap between training and analyzed recordings. (iv) Distances are in pixels without camera calibration, and the sampling rate, confidence threshold and image resolution were not varied. (v) The eight recordings share birds, compartments or sessions and are not independent flocks, so intervals describe within-recording uncertainty only. (vi) The optical-flow summary was selected after inspecting these recordings, and the region of interest was drawn after the primary results were known. Each limitation is detailed below.

Data leakage in model evaluation is a limitation, particularly for estimating prediction accuracy. The principal video contributed 532 of the 1528 training-pool frames (Table 1), and the training/validation split was frame-level for both models. Detection metrics (Table 2) are thus in-distribution and likely overestimate performance in a new house. The central rank-agreement result also replicated at $\rho = 0.77$–0.87 in four recordings with no training frames (Section 3.4), although three of them share birds or a session with training recordings (Table 1). Nonetheless, Table 2 should be read as a characterization of performance in these recordings, not a generalization.

Manual review establishes the provenance of the training annotations. Accuracy of predicted centroids and bird counts in held-out recordings remains unknown. Both missed and duplicate detections could contribute to count discrepancies. An independently annotated sample would permit measurement of count error, matched-centroid error in pixels and displacement error over verified trajectories.

All eight videos came from one laboratory at approximately ten birds per compartment. Four recordings supplied no development frames, but these recordings do not test transfer to other farms, stocking densities, housing systems or camera setups. No independently annotated locomotion or welfare outcome was available. The sampling rate, confidence threshold and image resolution were not varied in controlled sensitivity analyses. Differences in pixel scale, image sampling and retained detections can alter the displacement series. Time since housing ranged from 15 to 40 days and was not modeled; with ten birds per compartment the number of points per frame is unaffected, but body size and activity differ with age. Independently annotated trajectories, which would allow identity-preservation metrics such as the identification F1 score (IDF1) and the number of identity switches, were not available. Agreement with travel distance is therefore agreement between related computational measures, not validation against observed displacement.

4.4. Future Work

Future work should record groups of substantially different sizes to test whether the behavior of upper percentiles changes with the number of birds; subsampling $F(i)$ would not reproduce the detection errors of a larger flock. Another direction is to replace manual region cropping [1,8] with an automatic dynamic region of interest [42] that tracks pendular

equipment movement and defines the pen spatially with temporal hysteresis. This would restrict F(i) to the pen, excluding edge and equipment detections before Hausdorff calculation, and remove some optical-flow artefact. Further work includes a confidence-threshold and sampling-rate sweep, camera calibration for reporting in centimeters, and comparison with annotated trajectories and behavior.

5. Conclusions

In eight recordings of groups of ten broilers, deriving box centers and mask centroids from identical segmentation instances showed no consistent effect of the centroid definition on rank agreement with optical flow. The higher agreement of the segmentation pipeline came from the change of model, not from the centroid.

In this exploratory comparison, the 95th-percentile Hausdorff distance gave small gains over the maximum in all recordings, whereas the median lowered agreement. These gains do not establish practical superiority or an optimal percentile, and they apply only to recordings of about ten birds per frame.

Agreement with the detector-independent comparator, optical flow, was moderate ($\rho \approx 0.3$–0.6 on the full frame) and higher when flow was restricted to the pen floor. Agreement with tracker-derived travel distance was stronger but partly mechanical: about 0.19 of ρ came from shared coordinates, a lower bound because detections and identifiers are also shared.

Neither comparator provides independently observed displacement. Annotated trajectories, recordings from other houses and larger groups are needed before recommending a centroid definition or a Hausdorff percentile for estimating locomotion.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/ani16223653/s1: Table S1. Spearman correlation (ρ) of the classical Hausdorff Unrest Index (UI) and of tracker-derived travel distance (WD) with the four statistics of the block-mean optical-flow distribution (mean, variance, skewness, kurtosis), per recording. Point estimates over all valid processed steps; optical flow as in Table 3 (Farnebäck, frames at 0.5×, 8 × 8-pixel blocks); output file results/flow_statistics.csv. † No training frames. Table S2. Full-frame versus pen-floor region of interest (ROI), single-model pass, classical Hausdorff distance, and mask-area centroid. We kept instances when the mask centroid lay inside the pen-floor polygon or within 15 px of its edge; ROI flow is the mean of the 16 × 16-pixel blocks (half-resolution frame) whose center lies inside the polygon. $\Delta\rho = \rho(\text{ROI}) - \rho(\text{full frame})$ for the index–flow correlation; the centroid effect is mask − raw box center against ROI flow; and the model effect is segmenter − detector (both at the tracker box center) against ROI flow, with the detector index computed from all its detections. Paired moving-block bootstrap, 90 s blocks, 2000 replicates, seed 20261105; intervals describe uncertainty within each recording. † No training frames. Table S3. Stepwise correlations of Table 5 (principal recording) with 95% intervals from a moving-block bootstrap with 10 s blocks (2000 replicates, seed 0). Point estimates are identical to Table 5. Output file: results/correlations.json; command: src/analyze_correlations.py –block-sec 10 –nboot 2000 –seed 0 (the primary 90-s version: –block-sec 90, results/blocks90/correlations.json). Table S4. Block-length sensitivity of the controlled centroid contrast (mask − raw box center, classical Hausdorff distance), against optical flow and against own travel distance. Differences in Spearman ρ with nominal paired 95% bootstrap intervals (2000 replicates, seed 20261104); 90 s is the primary block length (Table 8). * Significant under the Holm procedure for the primary family within the recording. † No training frames. Output files: results/paired_centroids.json (10 and 90 s; –blocks 90,10) and results/long_blocks/paired_centroids.json (300 and 600 s; –blocks 300,600). For every block length, the model effect, the p95 − classical contrast, and the coordinate-read-out contrast stayed above zero in all eight recordings.

Author Contributions: Conceptualization: A.L.R.S.; methodology: A.L.R.S., C.N.H.F., M.M.N. and J.d.S.G.B.; software: A.L.R.S. and C.N.H.F.; formal analysis: A.L.R.S.; investigation: A.L.R.S. and I.C.H.; data curation: A.L.R.S. and I.C.H.; visualization: A.L.R.S.; writing—original draft: A.L.R.S.; validation: C.N.H.F. and J.d.S.G.B.; writing—review and editing: C.N.H.F., M.M.N., D.J.d.M. and J.d.S.G.B.; supervision: M.M.N. and D.J.d.M.; resources: D.J.d.M.; funding acquisition: D.J.d.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the São Paulo Research Foundation (FAPESP; grant 2024/21166-9) and the Brazilian National Council for Scientific and Technological Development (CNPq; grants 443961/2024-7 and 308177/2021-5).

Institutional Review Board Statement: The animal study protocol was approved by the Animal Use Ethics Committee of the University of Campinas (CEUA/UNICAMP), Brazil (protocol 6519-1/2024, approved on 21 August 2024; certificate 230/2024). The authorization for animal use was valid until 30 September 2026. This study analyzed existing recordings without additional handling or intervention in animals.

Informed Consent Statement: Written informed consent was obtained from the owner of the animals involved in this study.

Data Availability Statement: The analysis code, run configurations, aligned per-step time series, per-instance outputs and region-of-interest polygons are available at https://github.com/siriani/unrest-index-hausdorff-percentile (accessed on 30 August 2026) (release v2.2-revision). The repository README lists, for every table and supplementary table, the script, the exact command with its configuration (seed, block length, replicates) and the output file that contains the reported numbers. The eight recordings, the manually corrected annotations with the training and validation lists of both models, the per-instance outputs and the recording metadata are available at Zenodo, https://doi.org/10.5281/zenodo.23960761 (accessed on 30 August 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Del Valle, J.E.; Pereira, D.F.; Mollo Neto, M.; Gabriel Filho, L.R.A.; Salgado, D.D. Unrest index for estimating thermal comfort of poultry birds (Gallus gallus domesticus) using computer vision techniques. *Biosyst. Eng.* **2021**, *206*, 123–134. [CrossRef]
2. Pereira, D.F.; de Alencar Nääs, I. Estimating the thermoneutral zone for broiler breeders using behavioral analysis. *Comput. Electron. Agric.* **2008**, *62*, 2–7. [CrossRef]
3. Dawkins, M.S. Using behaviour to assess animal welfare. *Anim. Welf.* **2004**, *13*, S3–S7. [CrossRef]
4. Berckmans, D. General introduction to precision livestock farming. *Anim. Front.* **2017**, *7*, 6–11. [CrossRef]
5. Okinda, C.; Nyalala, I.; Korohou, T.; Okinda, C.; Wang, J.; Achieng, T.; Wamalwa, P.; Mang, T.; Shen, M. A review on computer vision systems in monitoring of poultry: A welfare perspective. *Artif. Intell. Agric.* **2020**, *4*, 184–208. [CrossRef]
6. Dawkins, M.S.; Cain, R.; Roberts, S.J. Optical flow, flock behaviour and chicken welfare. *Anim. Behav.* **2012**, *84*, 219–223. [CrossRef]
7. Dawkins, M.S.; Cain, R.; Merelie, K.; Roberts, S.J. In search of the behavioural correlates of optical flow patterns in the automated assessment of broiler chicken welfare. *Appl. Anim. Behav. Sci.* **2013**, *145*, 44–50. [CrossRef]
8. Pereira, D.F.; de Alencar Nääs, I.; Mehdizadeh, S.A. Equivalence between optical flow, the unrest index, and walking distance to estimate the welfare of broiler chickens. *Animals* **2025**, *15*, 1311. [CrossRef] [PubMed]
9. Pereira, D.F.; de Alencar Nääs, I.; Lima, N.D.d.S. Movement analysis to associate broiler walking ability with gait scoring. *AgriEngineering* **2021**, *3*, 394–402. [CrossRef]
10. Fernandes, A.M.; de Lucca Sartori, D.; de Oliveira Morais, F.J.; Salgado, D.D.; Pereira, D.F. Analysis of cluster and unrest behaviors of laying hens housed under different thermal conditions and light wave length. *Animals* **2021**, *11*, 2017. [CrossRef] [PubMed]
11. Huttenlocher, D.P.; Klanderman, G.A.; Rucklidge, W.J. Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **1993**, *15*, 850–863. [CrossRef]
12. Dubuisson, M.P.; Jain, A.K. A modified Hausdorff distance for object matching. In Proceedings of the 12th IAPR International Conference on Pattern Recognition, Jerusalem, Israel, 9–13 October 1994; pp. 566–568. [CrossRef]
13. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*, 29. [CrossRef] [PubMed]
14. Dawkins, M.S. Behaviour as a tool in the assessment of animal welfare. *Zoology* **2003**, *106*, 383–387. [CrossRef] [PubMed]
15. Ben Sassi, N.; Averós, X.; Estevez, I. Technology and poultry welfare. *Animals* **2016**, *6*, 62. [CrossRef] [PubMed]

16. Neethirajan, S. The role of sensors, big data and machine learning in modern animal farming. *Sens. Bio-Sens. Res.* **2020**, *29*, 100367. [CrossRef]
17. Wathes, C.M.; Kristensen, H.H.; Aerts, J.M.; Berckmans, D. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? *Comput. Electron. Agric.* **2008**, *64*, 2–10. [CrossRef]
18. Rowe, E.; Dawkins, M.S.; Gebhardt-Henrich, S.G. A systematic review of precision livestock farming in the poultry sector: Is technology focussed on improving bird welfare? *Animals* **2019**, *9*, 614. [CrossRef] [PubMed]
19. Li, N.; Ren, Z.; Li, D.; Zeng, L. Review: Automated techniques for monitoring the behaviour and welfare of broilers and laying hens: Towards the goal of precision livestock farming. *Animal* **2020**, *14*, 617–625. [CrossRef] [PubMed]
20. Dawkins, M.S.; Lee, H.J.; Waitt, C.D.; Roberts, S.J. Optical flow patterns in broiler chicken flocks as automated measures of behaviour and gait. *Appl. Anim. Behav. Sci.* **2009**, *119*, 203–209. [CrossRef]
21. Roberts, S.J.; Cain, R.; Dawkins, M.S. Prediction of welfare outcomes for broiler chickens using Bayesian regression on continuous optical flow data. *J. R. Soc. Interface* **2012**, *9*, 3436–3443. [CrossRef] [PubMed]
22. Colles, F.M.; Cain, R.J.; Nickson, T.; Smith, A.L.; Roberts, S.J.; Maiden, M.C.J.; Lunn, D.; Dawkins, M.S. Monitoring chicken flock behaviour provides early warning of infection by human pathogen campylobacter. *Proc. R. Soc. B* **2016**, *283*, 20152323. [CrossRef] [PubMed]
23. Gebhardt-Henrich, S.G.; Stratmann, A.; Dawkins, M.S. Groups and individuals: Optical flow patterns of broiler chicken flocks are correlated with the behavior of individual birds. *Animals* **2021**, *11*, 568. [CrossRef] [PubMed]
24. Kristensen, H.H.; Cornou, C. Automatic detection of deviations in activity levels in groups of broiler chickens—A pilot study. *Biosyst. Eng.* **2011**, *109*, 369–376. [CrossRef]
25. Aydin, A. Using 3D vision camera system to automatically assess the level of inactivity in broiler chickens. *Comput. Electron. Agric.* **2017**, *135*, 4–10. [CrossRef]
26. Farnebäck, G. Two-frame motion estimation based on polynomial expansion. In *Image Analysis (SCIA 2003)*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 363–370. [CrossRef] [PubMed]
27. Pereira, D.F.; Lopes, F.A.A.; Gabriel Filho, L.R.A.; Salgado, D.D.; Mollo Neto, M. Cluster index for estimating thermal poultry stress (Gallus gallus domesticus). *Comput. Electron. Agric.* **2020**, *177*, 105704. [CrossRef]
28. Massari, J.M.; de Moura, D.J.; de Alencar Nääs, I.; Pereira, D.F.; Branco, T. Computer-vision-based indexes for analyzing broiler response to rearing environment: A proof of concept. *Animals* **2022**, *12*, 846. [CrossRef] [PubMed]
29. Tan, X.; Yin, C.; Li, X.; Cai, M.; Chen, W.; Liu, Z.; Wang, J.; Han, Y. SY-Track: A tracking tool for measuring chicken flock activity level. *Comput. Electron. Agric.* **2024**, *217*, 108603. [CrossRef]
30. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [CrossRef]
31. Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLOv8 (Software, Version 8.0.0). 2023. Available online: https://github.com/ultralytics/ultralytics (accessed on 30 August 2026).
32. Jocher, G.; Qiu, J.; Liu, M.; Lyu, S.; Akyon, F.C.; Kalfaoglu, M.E. Ultralytics YOLO26: Unified real-time end-to-end vision models. *arXiv* **2026**, arXiv:2606.03748.
33. Sekachev, B.; Manovich, N.; Zhiltsov, M.; Zhavoronkov, A.; Kalinin, D.; Hoff, B.; Osmanov, T.; Kruchinin, D.; Zankevich, A.; Sidnev, D.; et al. *opencv/cvat: Computer Vision Annotation Tool (CVAT) (Version 2.73.0)*; Zenodo: Geneva, Switzerland, 2020. [CrossRef]
34. Neethirajan, S. ChickTrack—A quantitative tracking tool for measuring chicken activity. *Measurement* **2022**, *191*, 110819. [CrossRef]
35. Neethirajan, S. Automated tracking systems for the assessment of farmed poultry. *Animals* **2022**, *12*, 232. [CrossRef] [PubMed]
36. Siriani, A.L.R.; Kodaira, V.; Mehdizadeh, S.A.; de Alencar Nääs, I.; de Moura, D.J.; Pereira, D.F. Detection and tracking of chickens in low-light images using YOLO network and Kalman filter. *Neural Comput. Appl.* **2022**, *34*, 21987–21997. [CrossRef]
37. Siriani, A.L.R.; Miranda, I.B.C.; Mehdizadeh, S.A.; Pereira, D.F. Chicken tracking and individual bird activity monitoring using the BoT-SORT algorithm. *AgriEngineering* **2023**, *5*, 1677–1693. [CrossRef]
38. Zhang, Y.; Sun, P.; Jiang, Y.; Yu, D.; Weng, F.; Yuan, Z.; Luo, P.; Liu, W.; Wang, X. ByteTrack: Multi-object tracking by associating every detection box. In *Computer Vision—ECCV 2022*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 1–21. [CrossRef]
39. Doornweerd, J.E.; Veerkamp, R.F.; de Klerk, B.; van der Sluis, M.; Bouwman, A.C.; Ellen, E.D.; Kootstra, G. Tracking individual broilers on video in terms of time and distance. *Poult. Sci.* **2024**, *103*, 103185. [CrossRef] [PubMed]
40. Mehdizadeh, S.A.; Siriani, A.L.R.; Pereira, D.F. Optimizing deep learning algorithms for effective chicken tracking through image processing. *AgriEngineering* **2024**, *6*, 2749–2767. [CrossRef]
41. van der Eijk, J.A.J.; Guzhva, O.; Voss, A.; Möller, M.; Giersberg, M.F.; Jacobs, L.; de Jong, I.C. Seeing is caring—Automated assessment of resource use of broilers with computer vision techniques. *Front. Anim. Sci.* **2022**, *3*, 945534. [CrossRef]
42. Siriani, A.L.R.; Pereira, D.F.; Barros, J.d.S.G.; de Moura, D.J. DynROI-Hybrid: Dynamic modeling of functional regions for feeding and drinking behavior detection in laying hens. *Comput. Electron. Agric.* **2026**, *248*, 111757. [CrossRef]

43. Carion, N.; Gustafson, L.; Hu, Y.-T.; Debnath, S.; Hu, R.; Surís, D.; Ryali, C.K.; Alwala, K.V.; Khedr, H.; Huang, A. SAM 3: Segment Anything with Concepts (Model and Documentation). 2025. Available online: https://ai.meta.com/sam3/ (accessed on 30 August 2026).
44. Künsch, H.R. The jackknife and the bootstrap for general stationary observations. *Ann. Stat.* **1989**, *17*, 1217–1241. [CrossRef]
45. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.