CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Vision-based livestock pose estimation for precision livestock farming: A systematic review and design framework
Authors	Menolotto, Matteo;Tedesco, Salvatore;O'Grady, Luke;Kennedy, Emer;O'Flynn, Brendan
Publication date	2026-09-01
Original Citation	Menolotto, M, Tedesco, S, O'Grady, L, Kennedy, E & O'Flynn, B 2026, 'Vision-based livestock pose estimation for precision livestock farming: A systematic review and design framework', Computers and Electronics in Agriculture, vol. 251, 111951, pp. 1-32. https://doi.org/10.1016/j.compag.2026.111951
Type of publication	Review
Link to publisher's version	https://www.scopus.com/pages/publications/105041114041 - 10.1016/j.compag.2026.111951
Rights	© 2026, authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/) - https://creativecommons.org/licenses/by/4.0/
Download date	2026-10-23 21:14:50
Item downloaded from	https://hdl.handle.net/10468/18974

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Computers and Electronics in Agriculture 251 (2026) 111951

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Review Article

Vision-based livestock pose estimation for precision livestock farming: a systematic review and design framework

Matteo Menolotto [a,1,*], Salvatore Tedesco [a,b], Luke O'Grady [c], Emer Kennedy [d], Brendan O'Flynn [a]

[a] Tyndall National Institute, University College Cork, Lee Maltings Complex Dyke Parade, Cork, Ireland
[b] University College Cork, School of Computer Science and Information Technology, Western Rd, Mardyke, Cork, Ireland
[c] University College Dublin, School of Veterinary Medicine, Veterinary Science Centre Belfield Dublin 4, Ireland
[d] Teagasc, Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy Co. Cork, Ireland

ARTICLE INFO

Keywords:
Anatomical landmark detection
Lameness detection
Body measurement
Animal welfare monitoring
Multi-view imaging
On-farm monitoring

ABSTRACT

Camera-based pose estimation is becoming a key sensing approach for automated livestock monitoring, yet the selection of camera geometry, keypoint schema, learning model, and deployment strategy remains highly heterogeneous across species and applications. This systematic review synthesised the literature published or available online between 1 January 2015 and 31 December 2025 to identify task-aligned design patterns, reporting gaps, and practical routes toward on-farm implementation. Searches of ScienceDirect, Scopus, Web of Science, and Embase identified 273 records. After duplicate removal, eligibility screening, and methodological quality appraisal, 114 studies were included in the final synthesis.

The reviewed literature shows that cattle and pigs dominate the field, while poultry, horses, goats, sheep, camels, and cross-species studies remain comparatively under-represented. Across applications, side-view RGB is most consistently suited to gait and lameness analysis, whereas fixed top-down RGB offers the best trade-off for group-level behaviour, feeding, tracking, and identity-related tasks. Stereo, RGB-D, and multi-view systems are most justified when the target output depends on absolute geometry, persistent occlusion handling, or three-dimensional posture recovery, particularly in body measurement and weight-estimation pipelines.

The synthesis further shows that performance depends more on task-camera-keypoint alignment than on any single model family. However, progress toward robust deployment is limited by inconsistent landmark definitions, fragmented datasets, heterogeneous evaluation metrics, and insufficient reporting of runtime, calibration burden, and farm-system integration. A task-oriented reporting structure and preliminary species-specific core keypoint guidance are therefore proposed to improve comparability and reuse. Overall, livestock pose estimation is sufficiently mature for several structured farm applications, but broader adoption will depend on stronger standardisation, more diverse multi-site datasets, and more explicit validation under real operational conditions.

1. Introduction

Precision livestock farming (PLF) aims to transform livestock management from intermittent, labour-intensive observation into continuous, data-driven phenotyping of health, welfare, behaviour, and productivity (Berckmans, 2014). Non-contact sensing is particularly attractive in this context because it can reduce handling stress, operate at herd scale, and support earlier intervention than periodic manual inspection (Lovarelli et al., 2020; Romano et al., 2023). Among the available sensing modalities, camera-based systems have become especially relevant because they can capture posture, movement, body shape, and social context simultaneously from the same visual stream.

Within this visual ecosystem, pose estimation occupies a distinctive position. Rather than describing the animal only through whole-body silhouettes, bounding boxes, or global image features, pose estimation localises anatomically meaningful keypoints and preserves the spatial relations among them. This keypoint-based representation provides an interpretable intermediate layer for downstream inference of gait,

* Corresponding author.
E-mail address: matteo.menolotto@tyndall.ie (M. Menolotto).
[1] Present address: Tyndall National Institute, Lee Maltings Complex Dyke Parade, Cork, Cork T12 R5CP, Ireland.

https://doi.org/10.1016/j.compag.2026.111951
Received 23 January 2026; Received in revised form 21 May 2026; Accepted 23 May 2026
Available online 4 June 2026
0168-1699/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

posture transitions, feeding behaviour, social interaction, body dimensions, and disease-related anomalies (Jiang et al., 2022). It is therefore particularly suitable for PLF applications in which the mechanism of interest is biomechanical or postural rather than purely appearance-based.

At the same time, livestock pose estimation should not be considered in isolation from adjacent computer-vision paradigms. In practice, 2D RGB (red–green–blue) pose pipelines coexist with RGB-D (red–green–blue plus depth) cameras, stereo rigs, multi-view reconstruction, point-cloud analysis, Light Detection and Ranging (LiDAR), and Structure-from-Motion (SfM) photogrammetry. These alternatives are often more appropriate when the target variable depends on absolute three-dimensional geometry, such as body volume, linear morphometrics, or centimetre-level conformation traits, whereas 2D keypoint pipelines are often sufficient when the target is relative movement, posture, or repeated behavioural events (Wang et al., 2023a, 2023b; Dohmen et al., 2022; Ma et al., 2024). The critical design question is therefore not whether one modality is universally superior, but which sensing and modelling configuration is best matched to the biological trait, farm layout, and deployment constraints.

1.1. Pose estimation, 3D shape analysis, and main application areas in precision livestock farming

In the present review, livestock pose estimation refers to the detection and tracking of anatomically relevant landmarks (or keypoints) from visual data in order to reconstruct body configuration in two- or three-dimensional space (Jiang et al., 2022). The underlying pipelines may be top-down or bottom-up, single-frame or temporal, single-animal or multi-animal, and may rely on RGB, RGB-D, stereo, infrared, or multi-view inputs. By linking detected landmarks according to species-specific anatomical definitions, these systems generate interpretable spatial and temporal descriptors that can be propagated to downstream phenotyping tasks. Examples of general keypoint annotation schemes are shown in Fig. 1.

A first major application area is mobility and lameness assessment. In dairy cattle in particular, lameness remains a major welfare and productivity issue, yet conventional assessment is still largely based on subjective visual scoring of gait, stride regularity, back posture, and head movement (Thomsen et al., 2008; Schlageter-Tello et al., 2018). Pose estimation enables objective tracking of limb trajectories, joint motion, spinal curvature, and weight-shift asymmetries, thereby supporting earlier and more repeatable detection of locomotor abnormalities (Viazzi et al., 2014; Kang et al., 2021; Van Hertem et al., 2018). For this class of task, relative motion patterns and temporal consistency are often more important than absolute three-dimensional reconstruction (example in Fig. 2).

A second major area is growth monitoring, body dimension measurement, and live-weight estimation. In these applications, landmarks such as the withers, hips, tail base, chest, and back contour can be used to derive body length, height, width, or shape descriptors that correlate with growth and live weight (Wang et al., 2023a; Ma et al., 2024). Here, pose estimation frequently operates together with three-dimensional shape analysis: keypoints provide semantic anchors for region selection or measurement definition, while depth, stereo, LiDAR, or reconstructed point clouds improve metric fidelity and robustness to partial occlusion. Representative measurement pipelines are shown in Fig. 3.

A third area is body condition and welfare-related phenotyping. Accurate localisation of hooks, pins, tail head, spine, and related anatomical regions can support automated body condition scoring (BCS), while temporal pose trajectories can capture posture transitions, feeding events, restlessness, abnormal behaviours, and social interaction patterns (Zin et al., 2020; Lynn et al., 2017). In these settings, pose estimation is valuable not only because it improves automation, but also because it provides a more interpretable representation than end-to-end image classification. Representative lameness- and BCS-oriented examples are shown in Fig. 4.

Overall, pose estimation functions as a semantic bridge between raw visual sensing and biologically meaningful livestock phenotypes. Its practical value lies in the fact that the same keypoint representation can support multiple downstream tasks; however, its performance remains strongly conditioned by viewpoint, landmark visibility, annotation design, species morphology, and the degree to which three-dimensional information is required by the target phenotype.

1.2. Critical analysis of existing literature reviews

Several recent reviews are directly relevant to this topic, but none resolves the specific methodological choices required for pose-based livestock monitoring. Fazzari et al. (2025) survey deep-learning methods for animal behaviour analysis across multiple species and modalities, and distinguish pose-based from non-pose-based approaches, yet livestock-specific issues such as camera placement, anatomical annotation design, and farm deployment constraints are not their central focus. Rohan et al. (2024) provide a systematic review of deep learning for livestock behaviour recognition, but their synthesis is centred on behaviour classes and recognition models rather than on pose estimation as the sensing representation itself. Jiang et al. (2022) offer an important cross-species foundation for animal pose estimation, including datasets, two- and three-dimensional pipelines, and generic methodological challenges, but agricultural deployment conditions such as crowding, dirt, uncontrolled illumination, infrastructure variability, and edge hardware limitations receive limited treatment.

A related but distinct literature addresses three-dimensional

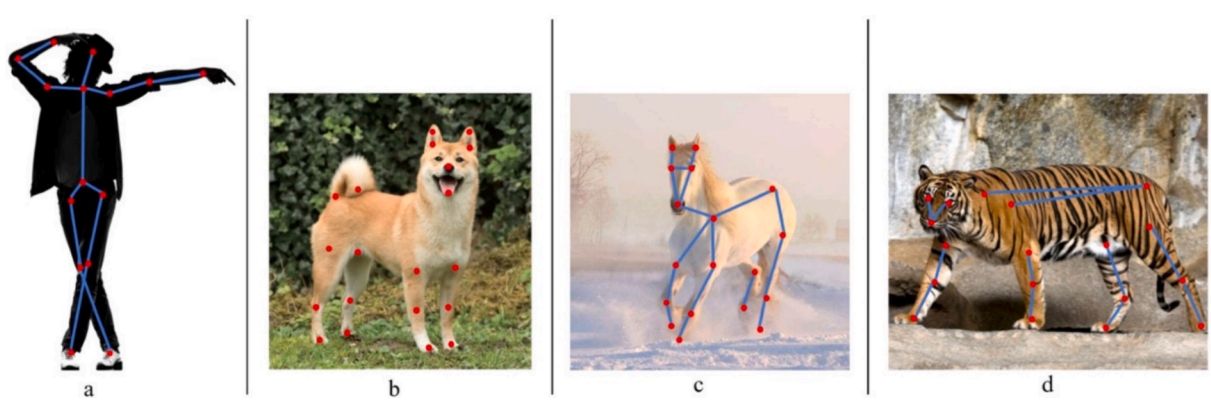

Fig. 1. Different keypoints and skeleton annotations on humans and animals, where (a) shows a standard for human pose annotation used in Andriluka et al. (2014), while three types of annotation for animal pose are from (b) Biggs et al. (2020), (c) Cao et al. (2019), and (d) Del Pero et al. (2017). The red points mark the keypoints on the image and the blue lines connecting the keypoints denote the skeleton.

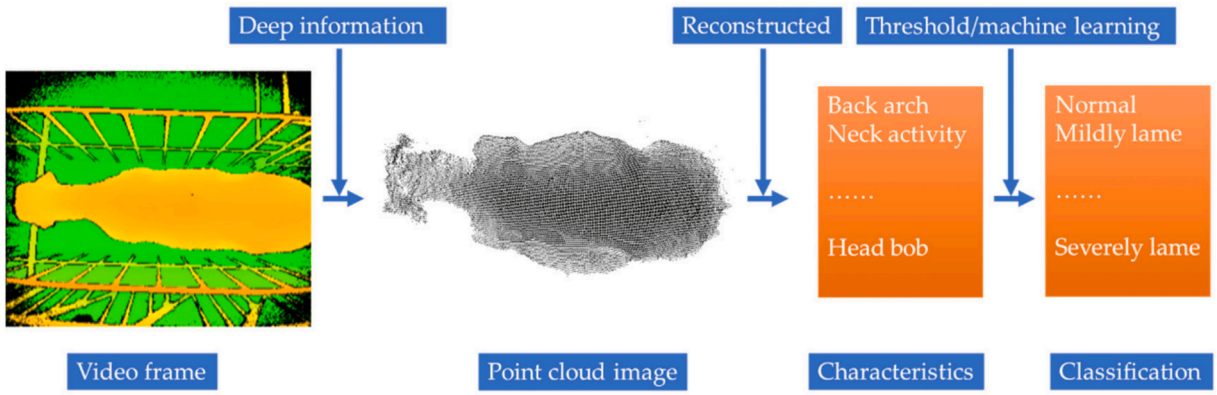

Fig. 2. Example workflow for lameness detection using a depth-based imaging pipeline, showing depth acquisition, point-cloud reconstruction, feature extraction, and downstream classification of locomotion status (adapted from Kang et al., 2021).

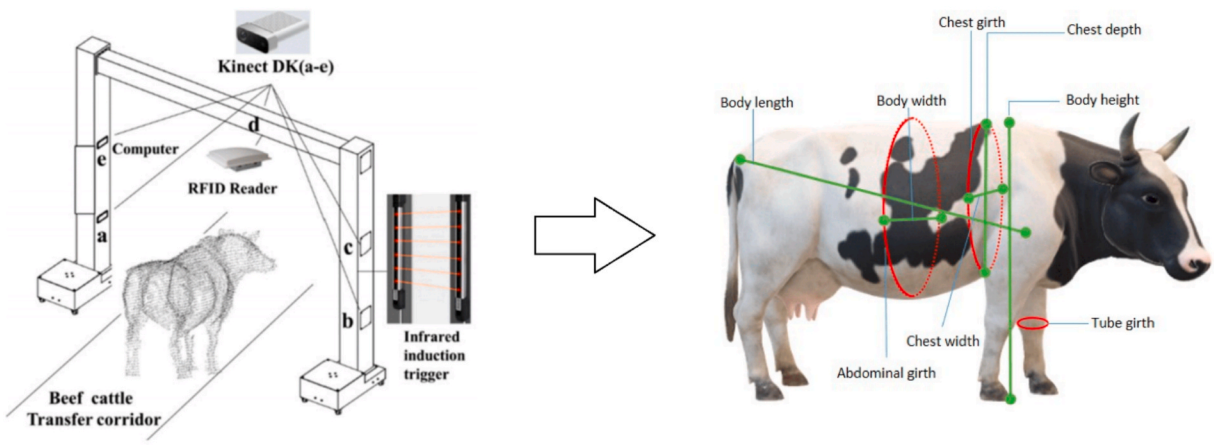

Fig. 3. Representative setups for automatic livestock body-dimension measurement: (left) channel-based acquisition with depth cameras and RFID-linked passage control; (right) example of standard morphometric traits derived from reconstructed 3D body volumes (adapted from Li et al., 2022a; Ma et al., 2024).

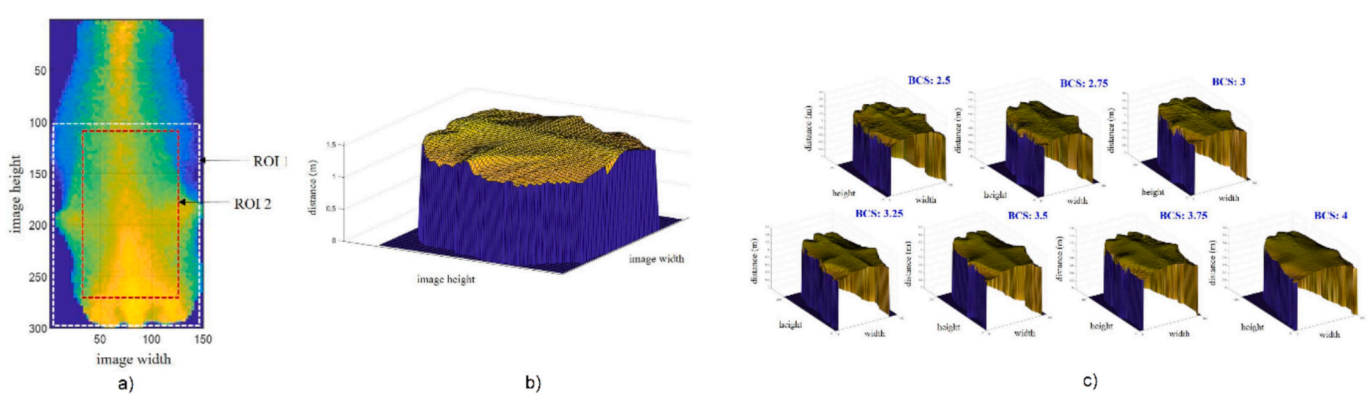

Fig. 4. Example of body condition score estimation from a 3D camera, showing (a) the depth image, (b) the reconstructed 3D body surface, and (c) cross-sectional regions of interest used for feature extraction in the scoring model (adapted from Zin et al., 2020).

livestock phenotyping. Wang et al. (2023a) review three-dimensional computer vision for cattle growth management, while Dohmen et al. (2022) and Ma et al. (2024) review visual approaches for body weight and body dimension measurement. These studies are valuable for understanding RGB-D, point-cloud, LiDAR, and reconstruction-based systems, but they do not systematically examine keypoint definition, landmark visibility, temporal pose modelling, or pose-driven downstream inference. Similarly, reviews of lameness detection and welfare monitoring often emphasise force plates, accelerometers, thermography, silhouettes, or global locomotion descriptors rather than anatomical landmark detection and tracking (Alsaaod et al., 2019; O'Leary et al., 2020).

Three gaps therefore remain. First, the literature still lacks a dedicated systematic review that treats pose estimation itself, rather than behaviour recognition or body measurement alone, as the central analytical layer in terrestrial livestock monitoring. Second, keypoint detection, three-dimensional shape analysis, and alternative visual technologies are usually discussed in parallel literatures, leaving limited

guidance on when RGB pose is sufficient and when RGB-D, stereo, LiDAR, or SfM are justified. Third, deployment-relevant factors such as camera geometry, annotation burden, cross-farm generalisation, latency, compute footprint, maintenance requirements, and species-specific anatomical definitions are seldom synthesised within a single design-oriented framework.

1.3. Aims and contribution of this review

To address these gaps, the present review synthesises the evidence on camera-based livestock pose estimation published between 2015 and 2025 and positions pose pipelines within the broader landscape of visual livestock phenotyping. The scope is restricted to terrestrial livestock studies in which anatomically relevant keypoints are inferred from image or video data. Non-visual-only systems, purely segmentation-based pipelines without a pose representation, and post-slaughter applications are outside the scope of the review.

The contributions of the review are fourfold. First, it maps the evolution of the field across species, tasks, camera geometries, and model families by using an updated corpus of 114 included studies. Second, it compares two-dimensional RGB pose pipelines with depth-enhanced, stereo, multi-view, and adjacent three-dimensional sensing strategies from a task-matching perspective rather than from an architecture-only perspective. Third, it synthesises how keypoints are defined across species and applications, highlighting the need for preliminary core landmark sets and more consistent reporting of annotation choices. Fourth, it derives a deployment-oriented design framework linking task, sensing configuration, keypoint schema, model class, evaluation metrics, and operational constraints such as throughput, latency, calibration effort, and housing conditions.

By focusing on the interaction between biological objective, sensing geometry, and deployment context, this review is intended to support both researchers developing new methods and practitioners selecting realistic on-farm configurations. The remainder of the manuscript is organised as follows. Section 2 describes the review protocol, search strategy, screening criteria, and data extraction process. Section 3 presents the results of the synthesis across the defined research questions. Section 4 discusses cross-cutting implications, standardisation needs, and future directions. Section 5 concludes the review with practical recommendations for research and deployment.

2. Methodology

2.1. Review protocol

This systematic review was conducted and reported in accordance with the PRISMA 2020 framework for evidence identification, screening, eligibility assessment, and inclusion (Page et al., 2021). The review also drew on the structured planning, conduct, and reporting logic described by Keele (2007) for systematic literature reviews in technical disciplines. The review protocol comprised four stages: (i) definition of the review scope and research questions; (ii) database search and record identification; (iii) duplicate removal, eligibility screening, and methodological quality assessment; and (iv) structured data extraction and narrative synthesis.

2.2. Research questions

The review was guided by nine research questions designed to characterise livestock pose-estimation systems from both a methodological and a deployment perspective. Collectively, these questions move from sensing and model design to data limitations, interpretability, species-specific constraints, and real-farm implementation.

RQ.1: What camera-based pose estimation techniques have been developed for livestock, and how do they perform in identifying key behaviours such as lying, grazing, walking, and social interaction under varied farm conditions?

RQ.2: What evidence exists on optimal camera hardware, placement, and viewing angles to ensure reliable pose estimation across diverse farm environments and livestock species?

RQ.3: How do depth-enhanced RGB systems and multi-camera setups compare to RGB-only approaches in terms of robustness to occlusions, lighting variability, and environmental complexity?

RQ.4: What machine learning architectures and training strategies (e.g., supervised, transfer learning, semi-supervised, or self-supervised approaches) have been most effective in extracting and interpreting livestock pose features for health and behaviour monitoring tasks?

RQ.5: What are the major barriers to creating large, diverse, and annotated livestock pose datasets, and how are methods such as synthetic data generation and domain adaptation being used to address them?

RQ.6: To what extent are explainable AI methods integrated into livestock pose-estimation pipelines, and how can interpretability be improved to support real-world decision-making by farmers and veterinarians?

RQ.7: How do environmental and species-specific factors, such as coat colour, animal density, body morphology, and housing systems, affect the accuracy and generalisability of pose-estimation systems?

RQ.8: What are the key deployment challenges, including computational load, power availability, connectivity, latency, calibration, and maintenance, when implementing pose-estimation systems in commercial farm settings?

RQ.9: How are keypoints defined across different livestock species, and what are the implications of anatomical variation for model design, training, and downstream behavioural or morphometric inference?

These selected research questions are mapped as follows:

- Pose Estimation Architectures and Performance: RQ.1, RQ.4
- Sensing Modalities and Camera Configuration: RQ.2, RQ.3
- Data Challenges and Solutions: RQ.5
- Interpretability and Trust: RQ.6
- Species-Specific Considerations (Anatomy, Keypoints, Behaviours): RQ.7, RQ.9
- Deployment Constraints and Real-World Performance: RQ.8

2.3. Databases and search strategy

For this study, four widely recognized databases were selected: ScienceDirect, Scopus, Web of Science, and Embase. The search strategy targeted peer-reviewed literature published, or available online ahead of final issue assignment, between 1 January 2015 and 31 December 2025, and combined keywords into database-specific search strings. Initially, the terms “pose detection” AND “livestock” were employed to identify relevant studies. To enhance the search results, the search string was refined to include synonymous terms, such as “pose prediction” and “posture detection” in place of “pose detection,” and “farm animals” and “animal husbandry” as alternatives for “livestock”. This refinement led to a general search string: (“pose detection” OR “pose estimation” OR “posture detection”) AND (“livestock” OR “farm animals” OR “cattle”) AND (“camera” OR “image-based”). Additional terms were included to cover specific farm animals like “goat” or “horse”, different camera-based technologies such as “depth” and “RGB” and to account for different synonyms, as summarised in Table 1.

ScienceDirect, Scopus, Web of Science and Embase were selected to provide broad multidisciplinary coverage across animal science, veterinary applications, agricultural engineering, computer vision, and precision-livestock-farming research. Scopus and Web of Science were used as broad citation databases to capture interdisciplinary engineering, sensing, and artificial-intelligence studies, while ScienceDirect provided additional coverage of Elsevier journals relevant to agricultural systems, biosystems engineering, animal science, and applied computer vision. Embase was included because pose-estimation

Table 1
Overview of criteria used as components of query on the four selected scientific literature databases. TS: the topic fields, including Title, Abstract and Keywords.

Criterion name	Description	Formula
Pose/ Keypoints	Pose estimation and keypoint/skeleton tracking terminology	TS = "pose estimation" OR "pose prediction" OR "posture detection" OR "keypoint detection" OR "skeleton tracking"
Livestock	Terrestrial farm animal species terms (livestock and common target species)	TS = "livestock" OR "cattle" OR "cow" OR "pig" OR "poultry" OR "chicken" OR "sheep" OR "goat" OR "camel" OR "horse" OR "farm animals" OR "animal husbandry"
Camera-based technology	Vision-based sensing and imaging modality terms	TS = "camera" OR "RGB" OR "depth" OR "vision-based" OR "image-based"

applications in livestock frequently intersect with animal health, welfare assessment, lameness detection, and veterinary decision support. IEEE Xplore and ACM Digital Library were not searched as separate primary sources. Although these databases are highly relevant to engineering, sensors, computer vision, and human–computer-interaction research, this decision may have limited the retrieval of some conference-oriented or engineering-specific studies. This is therefore acknowledged as a limitation of the search strategy.

2.4. Selection criteria

Eligibility was assessed using predefined inclusion and exclusion criteria aligned with the scope of the review. A study was retained only if it focused on terrestrial livestock and used image- or video-based methods to detect anatomically meaningful keypoints or skeleton structure for pose-related analysis. Studies were excluded if one or more of the following conditions applied:

1. the work focused exclusively on non-terrestrial animals, wildlife, or laboratory species;
2. pose estimation was not based on visual sensing;
3. no anatomically relevant keypoints were detected from image data;
4. the main objective was unrelated to pose estimation, for example pure behaviour classification, detection, or weight estimation without a pose representation;
5. the publication was a review, book chapter, conference abstract, editorial, thesis, or other non-primary source;
6. the article was not peer reviewed;
7. the full text was unavailable or not in English;
8. the record was a duplicate;
9. the work was published before 2015;
10. the study only applied proprietary commercial pose tools without methodological contribution; or
11. the study addressed only post-slaughter applications such as carcass processing or meat-quality grading.

The screening process followed the PRISMA sequence of identification, screening, eligibility, and inclusion. From the 273 records initially identified, 122 were retained after duplicate removal and preliminary screening against the scope and exclusion criteria. These 122 records then entered the methodological quality assessment stage.

2.5. Quality assessment

The 122 records retained after screening were assessed for methodological and reporting quality using six appraisal questions adapted from Kitchenham et al. (2009), which are suitable for structured review of computational and engineering studies. Each study received a score of 1 ("yes"), 0.5 ("partially"), or 0 ("no") for each criterion. The maximum total score was therefore 6. The quality assessment criteria were as follows:

1. Are the aims and objectives clearly stated?
2. Are the study scope, methodology, and experimental design clearly defined?
3. Is the research process documented appropriately?
4. Are the study questions addressed adequately?
5. Are limitations or negative findings reported?
6. Do the conclusions correspond to the study goals and presented results?

A minimum threshold of 3.0/6.0 was used to exclude papers with insufficient methodological transparency for reliable synthesis. On this basis, 114 studies were retained and 8 were excluded after quality appraisal. The final synthesis therefore reflects studies that met both the topical eligibility criteria and a minimum reporting-quality standard. Fig. 5 gives an overview of the PRISMA selection process.

2.6. Data extraction

The 114 included studies were then examined in detail using a structured data-extraction framework mapped to the review questions. The extracted variables covered species, behavioural or health endpoint, camera modality, camera placement, view geometry, number of animals in view, keypoint definitions, annotation approach, pose-estimation strategy, training regime, downstream models, evaluation metrics, deployment hardware, runtime indicators, limitations, and interpretability features.

In addition to conventional performance indicators, the revised extraction framework explicitly captured deployment-oriented variables such as inference speed, latency, computational footprint, calibration burden, edge-device compatibility, connectivity assumptions, and practical integration into farm workflows. This broader extraction strategy was adopted to support a more practice-oriented synthesis in the Results and Discussion sections.

The extracted information was synthesised narratively and comparatively rather than through *meta*-analysis because the reviewed studies were highly heterogeneous in species, tasks, keypoint schemas, datasets, and evaluation metrics. The final synthesis therefore emphasises cross-study patterns, design trade-offs, and reporting gaps rather than pooled effect sizes.

The synthesised data were used to address each research question and are presented in the following section (Table 2). An overview of the journals and conferences from which the selected studies were sourced, along with the geographical distribution of the research groups, is provided in Fig. 6 and Fig. 7, respectively.

3. Results

3.1. Current camera-based pose estimation methods for livestock, and how effective they are at recognising key behaviours in different farm environments (RQ.1)

This section presents the main descriptive characteristics of the included literature, with attention to species coverage, sensing setups, pose-estimation approaches, annotation ecosystems, and reporting practices. To contextualise the evidence base, Fig. 8 illustrates the distribution of livestock species across the included studies, with less frequent categories shown separately for clarity.

The uneven species distribution reflects more than publication bias. Cattle and pigs dominate partly because they combine high production value, well-defined welfare or productivity endpoints, and farming scenarios that are comparatively compatible with fixed-camera

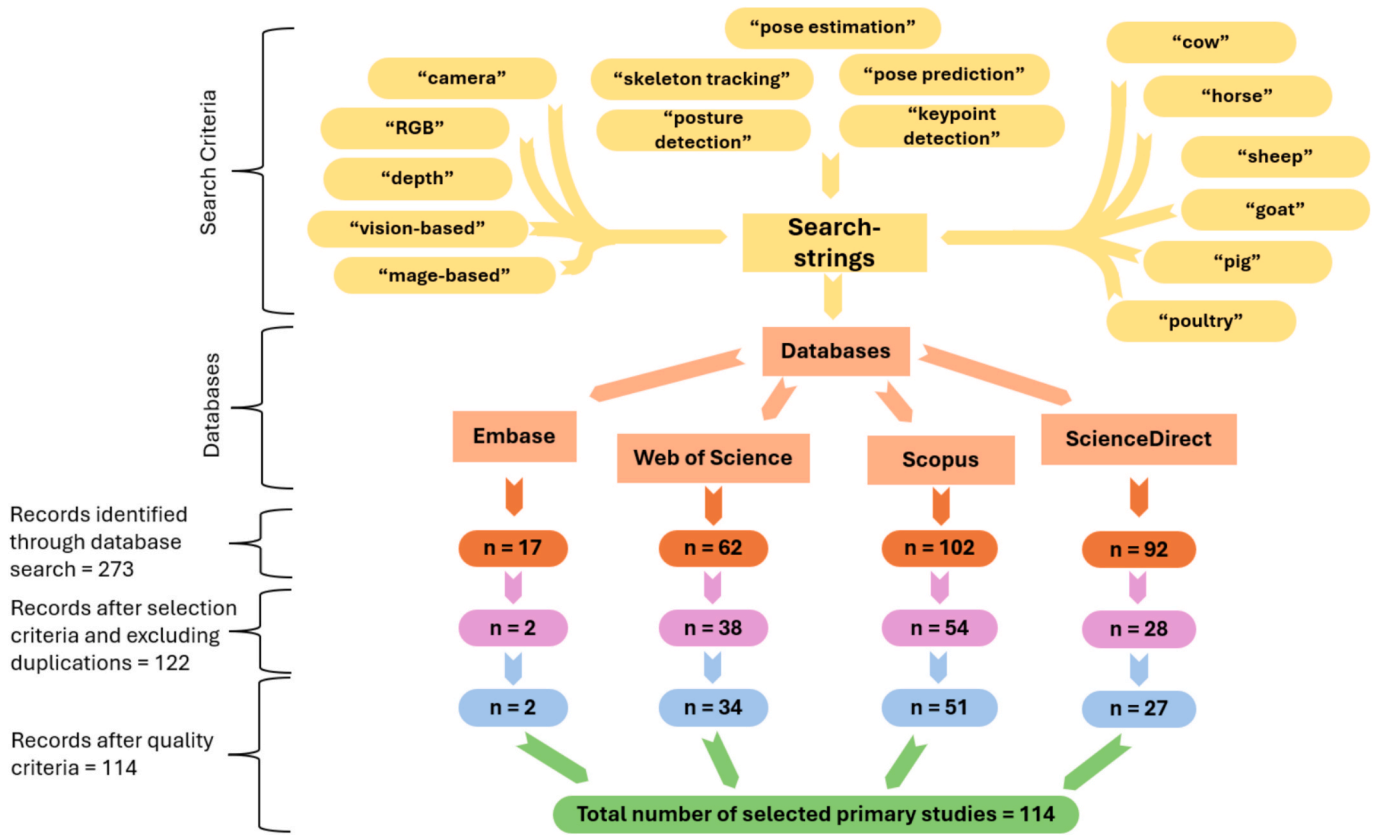

Fig. 5. PRISMA flowchart summarising the study selection process used in this review, from initial identification and screening to final inclusion of the eligible studies.

monitoring. In cattle, economically important traits such as lameness, body condition, live weight, growth, and identity can often be observed at repeated locations such as parlours, walkways, feeding areas, weighing corridors, or handling systems. These settings provide relatively stable viewpoints and facilitate data collection. In pigs, indoor group housing, feeding stations, farrowing pens, and nursery environments naturally support top-down or oblique camera installation, while welfare and behaviour endpoints such as posture, feeding, aggression, nursing, and social cohesion are directly relevant to production management.

Poultry, horses, sheep, goats, yaks, camels, and other livestock are less represented for different reasons. Poultry systems involve very high animal density, rapid motion, small distal anatomy, feather occlusion, and frequent overlap, making keypoint annotation and tracking difficult. Equine studies benefit from controlled clinical or breeding environments, but the number of animals and farms available for large-scale data collection is more limited. Sheep, goats, yaks, and camels are often managed in more extensive or less camera-constrained environments, with greater variation in coat, fleece, long hair, terrain, and husbandry practices. These differences also limit direct technical transfer from cattle or pig models. A pose schema, camera viewpoint, or pretrained model that works for top-down pig monitoring or side-view cattle lameness cannot be assumed to generalise to smaller ruminants, camelids, poultry, or extensive grazing systems without task- and species-specific validation.

3.1.1. State-of-the-Art Camera-Based pose estimation for livestock behaviour monitoring

Across the 114 included studies, three broad design families dominate: (i) top-down pipelines that first localise each animal and then estimate its keypoints; (ii) bottom-up pipelines that detect keypoints globally and subsequently assemble them into individual skeletons; and (iii) pose-informed downstream pipelines in which keypoints act as the intermediate representation for behaviour recognition, lameness scoring, morphometric estimation, tracking, or re-identification. Top-down approaches remain the most common overall because they are easier to train and often perform well when animal separation is feasible, whereas bottom-up methods are especially attractive in group-housed settings with overlapping animals.

Top-down approaches remain highly prevalent in pigs, poultry, horses, and controlled cattle corridors. Examples include broiler pose-based behaviour recognition using DeepLabCut or EfficientNet-based pipelines (Fig. 9) (Fang et al., 2021; Fang et al., 2024), sow and pig pose estimation with HRNet-, YOLOv8-, and transformer-derived backbones (Wang et al., 2022; Chen et al., 2025c; Liu et al., 2025a), dairy-cow gait and lameness analysis using T-LEAP, AiPEC, DeepLabCut, RTMPose, or Mask R-CNN (Russello et al., 2022; Taghavi et al., 2024; Higaki et al., 2024, 2025; Wu et al., 2025; Barney et al., 2023; Narli et al., 2025; Reitmaier and Narli, 2025), and horse kinematic or lameness monitoring with smartphone, stereo, or RGB-D systems (Feuser et al., 2022; Lawin et al., 2023; Niknejad et al., 2023; Li et al., 2025a, 2025b). In controlled or semi-controlled walkways, these pipelines often achieve high keypoint localisation performance and strong downstream classification accuracy.

Bottom-up or association-based pipelines are less numerous but remain important for herd scenarios. Representative examples include CMBN for multi-cow pose estimation (Fan et al., 2023), improved U-Net- and part-affinity-field-based cattle systems across multiple environments (Weng et al., 2023), multi-pig and multi-piglet pose/tracking systems using DeepLabCut, YOLOv8, BoT-SORT, or custom association logic (Farahnakian et al., 2021; Li et al., 2019; Kresovic et al., 2022; Chen et al., 2025a; He et al., 2025), and open-pasture cattle interaction recognition with pose graphs (Yang et al., 2023). Their main strength lies in better scalability to crowded scenes, although individual binding

Table 2
Mapping between research questions and data extraction fields.

RQ	Theme	Data extraction fields
RQ1	Architectures & performance	Species; behaviour/endpoint; single vs multi-animal; pose approach (top-down/bottom-up); tracking/temporal module; reported pose + downstream metrics; main failure modes.
RQ2	Camera configuration	Camera type (RGB/RGB-D/stereo/etc.); placement (top-down/side/oblique); distance/height; resolution/FPS; lighting constraints; number of animals in view; chokepoint vs pen-scale.
RQ3	Modalities comparison	RGB vs RGB-D/stereo vs multiview; occlusion strategy; calibration needs; robustness claims + evidence (conditions tested); metric error improvements (if any).
RQ4	Model architectures & training	Backbone/head family; temporal model; training regime (supervised/transfer/semi-supervised/self-training); augmentations; pretraining source; compute footprint (params/FLOPs/latency).
RQ5	Data bottlenecks & solutions	Dataset size; #farms/#sites; open access vs private; annotation protocol/tool; label noise mitigation; synthetic data; domain adaptation; cross-domain testing.
RQ6	Interpretability & trust	Explanation type (skeleton overlays, feature importances, saliency, 3D evidence); user-facing outputs; evidence of end-user evaluation; failure reporting.
RQ7	Species/environment factors	Breed/coat pattern issues; density/occlusion regime; housing type; background/lighting; keypoint visibility issues; generalisation notes.
RQ8	Deployment constraints	Inference hardware (edge/server); FPS/latency; connectivity/cloud dependence; calibration/maintenance; ID continuity; uptime robustness; operational key performance indicators.
RQ9	Keypoints & anatomy	Keypoint taxonomy (# and definitions); anatomical rationale; mesh alignment (if any); task linkage; implications for architecture choice and behavioural inference.

and distal-joint stability remain difficult under heavy occlusion.

A third, increasingly important category is the pose-informed multi-task or task-specific pipeline. These systems do not stop at keypoint localisation but propagate pose into downstream tasks such as posture classification, body measurement, weight estimation, social interaction analysis, rumination monitoring, aggression detection, brush-use quantification, identity recognition, or fever screening. Examples include MT-SRNet for simultaneous pig keypoints, masks, and posture (Liu et al., 2025a), pig feeding recognition from pose-derived spatial cues (Hu et al., 2025), goat aggression detection from tracked keypoint dynamics (Qu et al., 2025), dairy-cow brush-use analytics from keypoint-to-brush geometry (Högberg et al., 2025), cattle re-identification with pose-aware transformers (Liu et al., 2025b), and non-contact yak or cattle weight estimation using keypoint-driven geometric features (Wang et al., 2025; Bai et al., 2025; Rozendo et al., 2025) (Fig. 10). This confirms that pose estimation is increasingly being used as an interpretable substrate rather than as an endpoint in itself.

Temporal modelling is especially prominent in mobility and lameness applications. T-LEAP demonstrated that spatio-temporal convolution substantially improves robustness to occlusion and unseen animals in walking-cow videos (Russello et al., 2022), while later work adapted temporal pose methods to indoor housing and automated lameness classification (Taghavi et al., 2024; Russello et al., 2024). Sequence-aware classifiers also appear in poultry and cattle, including CNN-LSTM models for broiler gait (Nasiri et al., 2022), motion-feature-based lameness classifiers in cattle (Zhao et al., 2023a; Narli et al., 2025; Reitmaier and Narli, 2025; Wu et al., 2025), and temporal tracking refinements such as KeySORT or Kalman-based association for multi-animal continuity (Perneel et al., 2025a, 2025b; Li et al., 2019). Overall, the literature suggests that temporal information is most valuable when the phenotype of interest is fundamentally dynamic, such as lameness, gait, posture transitions, or aggressive interaction.

Graph-based and transformer-based methods are also becoming more common, but their value is task-dependent rather than universal. GCN (Graph Convolutional Network)-based coordinate refinement improved real-time pose estimation in dairy goats (Fig. 11) (Chen et al., 2024), graph reasoning supported cattle interaction recognition (Yang et al., 2023), pose-guided transformer attention improved cattle re-identification under viewpoint variation (Liu et al., 2025b), and enhanced transformer or attention modules were repeatedly used to stabilise keypoint detection under clutter, long hair, cage structures, or multi-scale anatomy (Wang et al., 2022; Zhao et al., 2024; Liu et al., 2025b; Xu et al., 2025a). However, the updated corpus still indicates that success depends more on task-camera-keypoint alignment than on any single backbone family.

Tables 3–7 summarise the most prominent pose-estimation strategies and the reporting landscape, spanning pose-localisation metrics, downstream task metrics, computational-efficiency indicators, and their relationship to livestock species and pose-detection strategies.

3.1.2. Reported performance and endpoint dependence

Performance must be interpreted relative to the downstream endpoint. For coarse posture recognition under controlled conditions, very high performance is common. Examples include broiler gait or posture recognition above 95% accuracy in structured settings (Nasiri et al., 2022), sow nursing detection above 97% accuracy (Gan et al., 2021), pig feeding recognition with near-perfect recall for feeding events (Hu et al., 2025), and goat aggression recognition at 95.04% accuracy with explicit false-positive control (Qu et al., 2025). These results show that when the camera view is stable, the keypoint schema is tightly linked to the behaviour, and class boundaries are clear, pose-informed systems can achieve strong reliability.

In contrast, more clinically nuanced or ecologically variable tasks remain harder. Mild-versus-moderate lameness remains difficult in both cattle and horses, even when overall binary detection is strong (Narli et al., 2025; Reitmaier and Narli, 2025; Higaki et al., 2024; Feuser et al., 2022). Similarly, brush-use segment classification, pig social cohesion, and multi-animal tracking in dense scenes remain sensitive to occlusion, viewpoint, and identity continuity (Högberg et al., 2025; He et al., 2025; Bibinbe et al., 2025). The literature therefore supports a practical distinction between tasks that are pose-separable under modest sensing assumptions and tasks that remain bottlenecked by occlusion, ambiguity, or annotation design.

The 2025 additions strengthen this interpretation. Narli et al. (2025) reported 97% accuracy for binary lame-versus-non-lame classification from 12 back keypoints, but performance dropped for adjacent locomotion-score strata. Reitmaier and Narli (2025) similarly showed that hierarchical lameness classification improved overall robustness but remained limited by the ambiguity of mild cases and by inter-observer disagreement. Wu et al. (2025) achieved strong pose and lameness performance in multi-cow scenes, yet severe occlusion still degraded the system. These results reinforce that high binary performance does not automatically translate into reliable fine-grained severity grading.

3.1.3. Evaluation metrics and a task-aligned recommendation set

A major obstacle to cross-study comparison remains metric heterogeneity. The updated corpus confirms that pose localisation is reported using a mixture of AP/mAP, OKS (Object Keypoint Similarity), PCK/PCKh (Percentage of Correct Keypoints / head-normalised Percentage of Correct Keypoints), RMSE (root mean square error), pixel error, and less standardised indicators, while downstream tasks add accuracy, precision, recall, F1-score, AUC (area under the curve), MAE (mean absolute error), MAPE (mean absolute percentage error), R^2, temporal consistency, tracking metrics, or application-specific agreement measures. This fragmentation makes it difficult to compare systems across papers, even when they appear to solve similar problems.

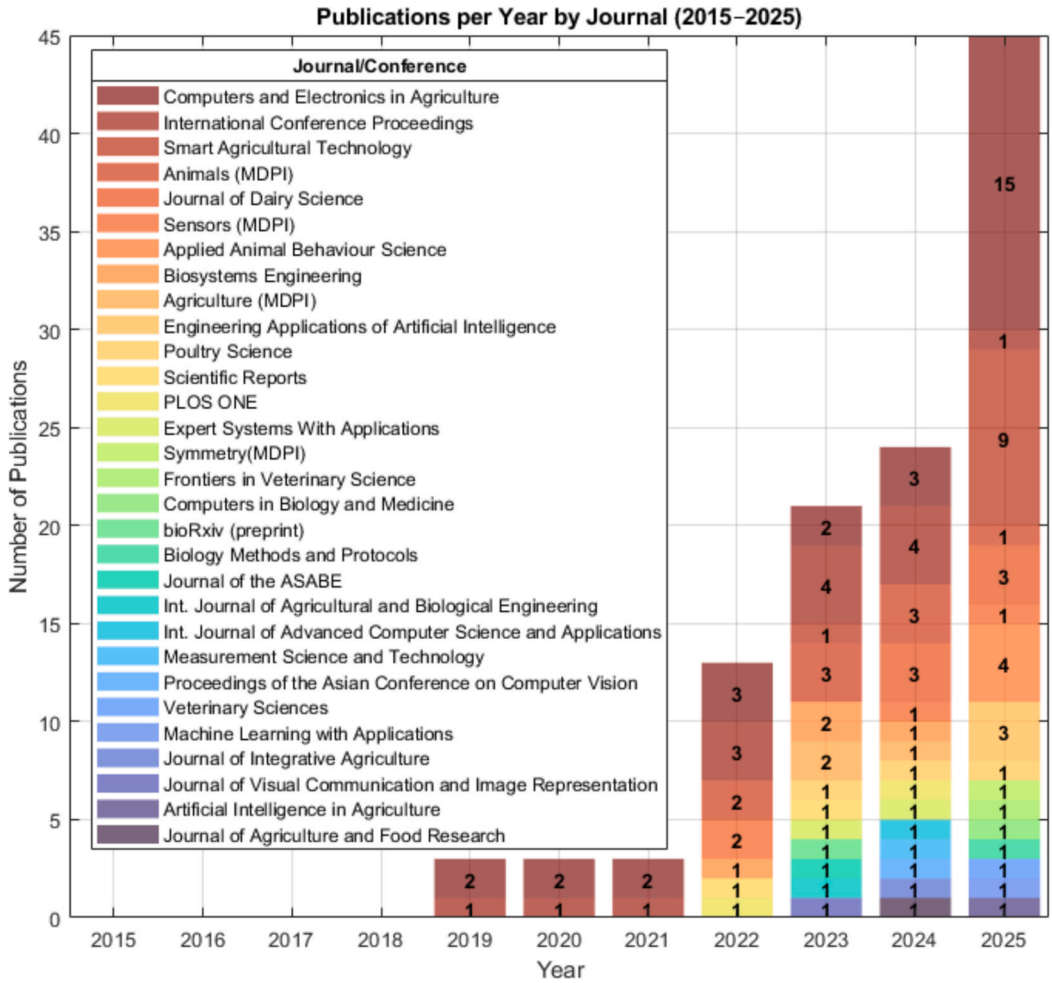

Fig. 6. The distribution of the selected primary studies by year and journal of publication.

The literature nevertheless suggests a clearer minimum reporting convention. For generic 2D/3D keypoint localisation, studies should report either OKS-based AP together with the sigma/visibility settings used, or PCK/PCKh together with the normalisation rule, plus a physical or pixel-space localisation error where relevant. For gait and lameness applications, localisation performance should be paired with downstream balanced accuracy or macro-F1, sensitivity/specificity, and a temporal stability measure, because clinically useful systems depend on sequence consistency rather than frame-wise accuracy alone. For morphometrics and weight estimation, physical-unit errors such as MAE, RMSE, and MAPE should be reported together with agreement analyses against manual measurements. For tracking and re-identification, identification F1 Score (IDF1), multiple object tracking accuracy (MOTA) or higher order tracking accuracy (HOTA), together with Rank-1/mAP where relevant, should accompany the pose metrics.

Accordingly, a compact task-aligned metric bundle is recommended: (i) behaviour and posture recognition: pose AP or PCK + macro-F1 + per-class recall; (ii) gait and lameness: pose PCKh or OKS + balanced accuracy/macro-F1 + sensitivity/specificity + temporal consistency; (iii) body measurement and weight estimation: pose AP/OKS + MAE/RMSE/MAPE in physical units + agreement statistics; and (iv) multi-animal tracking or re-identification: pose AP/OKS + IDF1/MOTA/HOTA or Rank-1/mAP. This recommendation goes beyond merely identifying heterogeneity and provides a practical reference for future reporting.

3.1.4. Computational efficiency and representative edge performance

Only a subset of studies reported runtime or hardware results in sufficient detail, but the updated evidence base is stronger than in the previous version. Lightweight or edge-oriented systems now span cattle, pigs, poultry, goats, horses, and yaks. Examples include 39 FPS for improved RTMPose in cattle (Li et al., 2023), 28.3 FPS on Jetson Xavier NX for YOLOv8-DMC cattle measurement (Weng et al., 2025), around 24 FPS for real-time goat aggression analysis from pose and tracking (Qu et al., 2025), 24–27 FPS for pose and segmentation modules in the portable horse RGB-D device on Jetson Nano (Li et al., 2025b), and 16 FPS on Jetson Orin Nano for duck cage-net removal plus pose estimation (Zhao et al., 2025a).

However, edge readiness is highly model- and task-dependent. Some systems remain too slow for true continuous video deployment on low-power devices, such as the yak-heifer model running at approximately 248 ms per frame on Jetson Nano, or the yak weight-estimation pipeline requiring around 0.78 s per inference on Raspberry Pi for the lightweight variant and several seconds for heavier variants (Li et al., 2025b; Wang et al., 2025). Similarly, full weight-estimation or point-cloud pipelines may be accurate but remain workstation- or GPU-oriented rather than field-real-time (Tan et al., 2025; Rozendo et al., 2025; Wei et al., 2025). The field is therefore progressing toward embedded deployment, but claims of "real-time" still require careful inspection of hardware, frame size, and total pipeline latency.

Together, Tables 5–7 show that the choice of evaluation metric is driven not only by model family, but also by livestock species, endpoint type, and whether the pipeline is aimed at localisation, downstream inference, or deployment readiness.

Overall, RQ.1 indicates that livestock pose estimation is no longer limited to proof-of-concept 2D keypoint detection. The field now

Fig. 7. Global distribution of the publications included in the review by country.

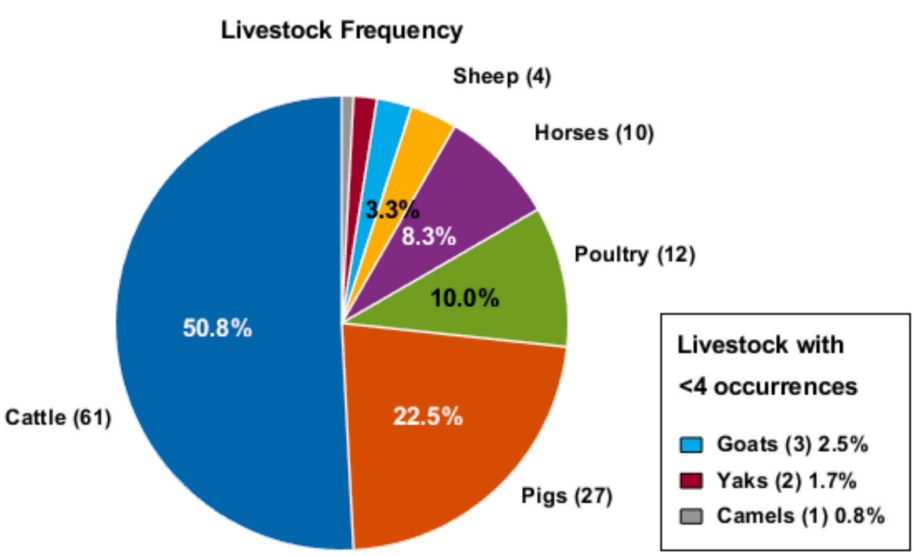

Fig. 8. Distribution of livestock species across the included studies. Less frequent categories are listed separately for clarity.

includes mature top-down pipelines for constrained farm tasks, a smaller but important set of bottom-up or association-based methods for group scenes, and a rapidly growing class of pose-informed systems for welfare, health, morphometrics, and identity analytics. The strongest current systems are those in which the camera geometry, keypoint schema, and downstream task are explicitly co-designed.

3.2. Optimal camera hardware and placement for livestock pose estimation (RQ.2)

3.2.1. Camera modalities and practical trade-offs

Commodity RGB remains the dominant sensing modality because it is inexpensive, flexible, and compatible with the largest ecosystem of pose-estimation toolchains (Fig. 12). RGB systems are particularly effective when the target trait depends on relative posture or motion

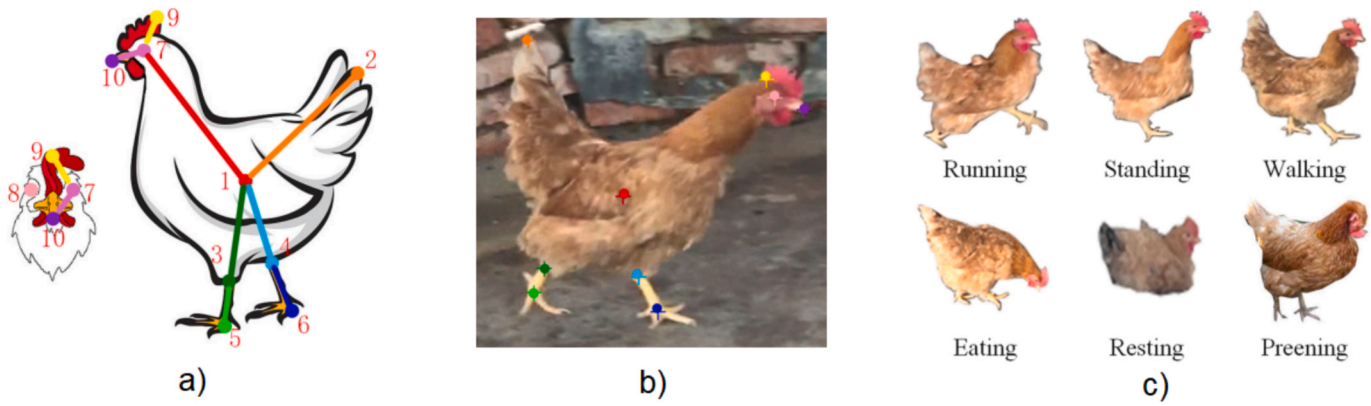

Fig. 9. (a) schematic representation of annotated keypoints on a broiler chicken; (b) example frame with detected body markers; (c) representative behaviour classes used to train the deep neural network model (Fang et al., 2021).

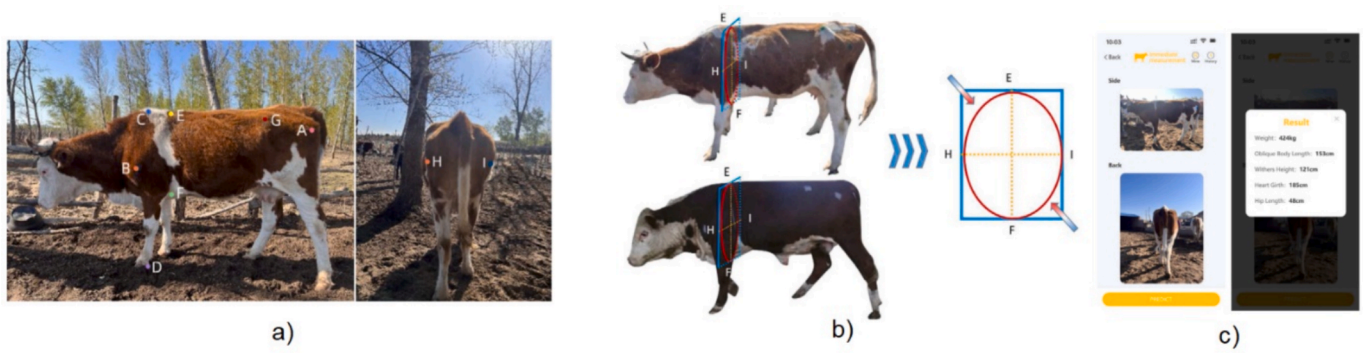

Fig. 10. A) keypoints annotated on the side and back views; b) chest width (ef) and hip width (hi) derived from these keypoints and used to compute ramanujan's ellipse underlying the weight estimation model; c) example of the mobile phone interface display (Bai et al., 2025).

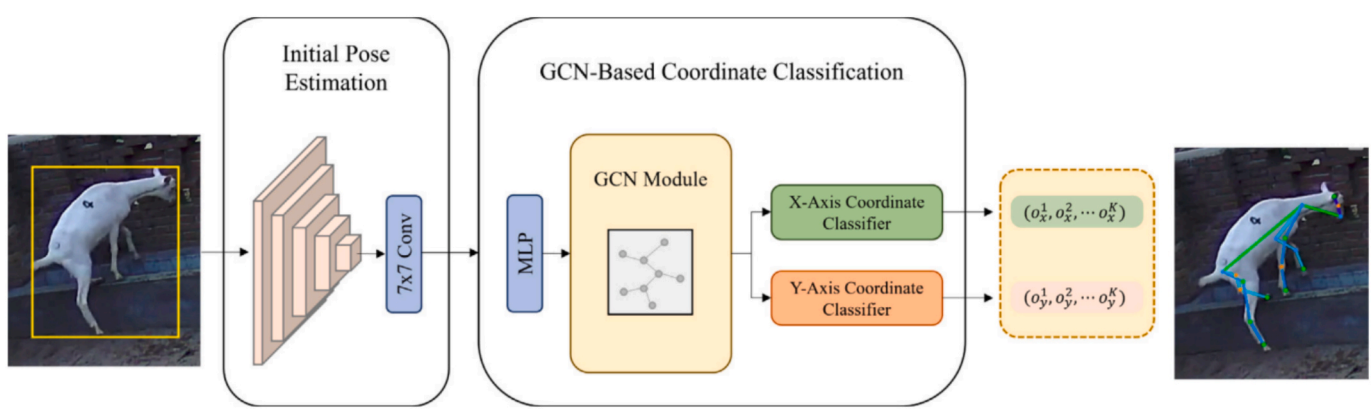

Fig. 11. Architecture of GRMPose, showing the two-stage dairy-goat pose-estimation pipeline (Chen et al., 2024).

rather than on absolute geometry. This is evident in many side-view lameness systems for cattle and horses (Russello et al., 2022; Taghavi et al., 2024; Higaki et al., 2024; Narli et al., 2025; Reitmaier and Narli, 2025; Lawin et al., 2023), top-down welfare or behaviour systems in pigs and goats (Gan et al., 2021; Wang et al., 2023b; Hu et al., 2025; Qu et al., 2025), and research-scale applications such as brush-use monitoring and cattle re-identification (Högberg et al., 2025; Menezes et al., 2025; Liu et al., 2025b).

RGB-D, stereo, or multi-view systems become more advantageous when the phenotype of interest depends on absolute body dimensions, depth ordering, or occlusion recovery. This is especially clear in cattle, pig, sheep, horse, yak, and poultry measurement studies in which 2D keypoints are back-projected to 3D or coupled to point clouds and consistently yield low relative errors for linear traits under controlled geometry (Du et al., 2022; Luo et al., 2023; Yang et al., 2024; Peng et al., 2024; Deng et al., 2025; Yue et al., 2025; Zhao et al., 2025b; Wei et al., 2025; Li et al., 2025a; Wang et al., 2025; Xu et al., 2025a). In these systems, the extra sensing complexity is justified because the endpoint itself is metric.

Multimodal extensions are used more selectively (Table 8). RGB plus optical flow, contour streams, infrared, or thermal imaging is mainly employed when the primary challenge is not measurement but robustness under poor lighting, cage structures, or subtle health signatures. Examples include RGB-optical-flow-skeleton fusion for cattle lameness

Table 3
Representative examples of prominent pose-estimation strategies.

Study	Strategy	Behaviour / Task	Performance Metric
Fang et al. (2022) – BroilerPose	Top-down with ResNet–50 + FPN	Posture (broiler chicken)	F1–score 89%, mAP 0.8652
Taghavi et al. (2024) – T–LEAP (cows)	Temporal keypoint detection	Gait/occlusion handling	PCKh@0.2: 89%; Precision 94.7%; Recall 87%
Chen et al. (2024) – GCN–based (goats)	Top-down keypoints + GCN classifier	Grazing/lying	AP: 87.5%; Latency: 13.8 ms
Fan et al. (2023) – CMBN (cattle)	Bottom-up for multi-cow tracking	Posture detection	AP: 93.2%
Barney et al. (2023) – Modified Mask–RCNN (cattle)	Detection with pose + SVM for lameness	Lameness classification	Accuracy: 94%

SVM= Support Vector Machine, mAP = mean average precision, AP = Average Precision.

Table 4
List and frequency of usage of object detection and keypoints evaluation metrics.

Evaluation Metric	Description	No of Works
Precision / AP / mAP	Overlap- or threshold-based detection/localisation performance, including AP-family variants (for example AP, mAP, AP50, AP75, AP90, AP_kp, AP_box).	51
OKS	Object Keypoint Similarity; keypoint localisation agreement normalised by object scale and keypoint uncertainty.	18
RMSE	Root Mean Square Error of landmark localisation or derived measurement error, typically in pixels or physical units.	14
PCK / PCKh	Percentage of Correct Keypoints, often normalised by body or head size.	11
Pixel Error (PE)	Direct pixel-distance error between predicted and annotated landmarks.	5

Table 5
List and frequency of usage of classification and regression metrics.

Evaluation Metric	Description	No. of Works
Accuracy	Overall proportion of correct predictions.	26
Recall / AR / mAR	Recovery of true positives or annotated targets; includes class recall and average recall variants.	35
F1-score	Harmonic mean of precision and recall.	16
Sensitivity	True positive rate, commonly used in screening/ classification tasks.	9
Specificity	True negative rate, commonly used in screening/ classification tasks.	8
AUC (incl. AUC-ROC)	Threshold-independent discrimination performance.	5
Pearson Correlation	Correlation/agreement between automated outputs and reference measurements or scores; in the supplied studies this sometimes appears as Pearson or Spearman correlation.	3
Cohen's & Fleiss' Kappa	Agreement beyond chance between automated and/or human ratings.	4
Success Rate (SR)	Proportion of successful detections/localisations under a defined acceptance criterion.	2

(Li et al., 2022), RGB-thermal facial landmarking and fever estimation in cattle (Pham et al., 2025), cage-net elimination and image completion before duck pose estimation (Zhao et al., 2025a), and multimodal contour-skeleton fusion for sow lameness (Chen et al., 2025c). These studies show that alternative modalities are valuable, but usually as task-specific augmentations rather than general replacements for RGB.

Table 6
List and frequency of usage of computational efficiency metrics.

Evaluation Metric	Description	No. of Works
Model Parameters (incl. model size)	Parameter count and/or model footprint, reflecting deployment burden.	7
Detection Speed (FPS / inference speed)	Throughput of the pipeline, usually reported as frames per second or explicit inference speed.	6
FLOPs / G/MFLOPs	Computational complexity of the model.	5
Temporal Consistency (TC)	Frame-to-frame stability/continuity of structural predictions.	2
Variance / Standard Deviation (SD/STD)	Dispersion or stability of the reported performance/error.	1
Inference Latency	Time required to process a frame or sample.	1
Valid Cow Percentage (VCP)	Proportion of frames/animals producing a valid full-body structural output.	1

3.2.2. *Placement: top-down, side-view, oblique, and multi-view*

Camera placement is one of the strongest determinants of achievable performance. Top-down placement is favoured for counting, tracking, pen-scale activity analysis, feeding, nursing, cohesion, identity, and group-level interaction because it reduces self-occlusion, simplifies individual separation, and often allows one camera to monitor a larger area. This is seen repeatedly in pig and poultry work (Gan et al., 2021; Gorssen et al., 2022; Wang et al., 2022; Witte et al., 2024; Hu et al., 2025; Liu et al., 2025a; He et al., 2025), in cattle identity and mobility work (Menezes et al., 2025; Higaki et al., 2025; Ong et al., 2023; Zhang et al., 2025), and in goat aggression monitoring (Qu et al., 2025). Representative examples of top-down livestock imaging configurations used for keypoint detection and identity-related monitoring are shown in Fig. 13.

Side-view placement remains preferable for gait, locomotion asymmetry, stride timing, head bobbing, back curvature, and many linear body measurements because sagittal kinematics are directly visible in that geometry. This is consistently observed in cattle lameness studies (Russello et al., 2022; Taghavi et al., 2024; Russello et al., 2024; Narli et al., 2025; Reitmaier and Narli, 2025; Wu et al., 2025; Zhao et al., 2023a), in equine lameness and gait analysis (Feuser et al., 2022; Lawin et al., 2023; Niknejad et al., 2023), and in many side-view cattle or horse morphometric systems (Bai et al., 2025; Li et al., 2025b; Rozendo et al., 2025; Weng et al., 2025).

Oblique or mixed-angle setups are typically compromises driven by barn architecture. They are sometimes advantageous when a pure top-down or pure side view is infeasible, but they also increase perspective distortion and complicate calibration. In practice, the literature suggests that oblique views work best when paired with either strong post-processing or additional modalities. This is why many practically successful systems either constrain the animal path in corridors and walkways or move to multi-view sensing when the phenotype is sensitive to occlusion or depth ambiguity.

Multi-view systems are justified when a single view is structurally insufficient. Seven-camera dairy-cow systems achieved strong timing accuracy for posture-transition analysis in freestall barns (Kroese et al., 2024a, 2025), four-camera fusion supported bird's-eye-view location mapping and behaviour monitoring in cattle barns (Nasir et al., 2025), and multi-camera datasets such as HanwooReID and BECA show the growing importance of pose-aware identity continuity across viewpoints (Liu et al., 2025b; Zhang et al., 2025a). The literature therefore supports a simple rule: single-view systems are often adequate when the path, view, and task are constrained; multi-view systems are preferable when occlusion, identity continuity, or 3D posture must be resolved under less controlled conditions.

Table 7
Relationship between evaluation metric families, livestock targeted, and pose-detection strategy in the reviewed studies.

Evaluation Metric Family	Livestock Most Commonly Targeted in the Supplied Studies	Typical Pose Detection Strategy / Reporting Context
Precision / AP / mAP family	Mainly cattle, then pigs and poultry	Most common in top-down/ direct keypoint localisation pipelines, but also frequent in bottom-up multi-animal and single-stage pose models. Usually used to benchmark keypoint detection or combined detection-pose systems.
OKS	Mainly cattle and pigs, with additional poultry use	Most strongly associated with pose localisation benchmarking, especially in top-down, bottom-up/PAF-based, and single-stage keypoint pipelines.
RMSE / Pixel Error	Mainly pigs and cattle, with some poultry and horse use	Common in DeepLabCut-style landmark regression, point-cloud/3D measurement pipelines, and studies reporting direct localisation error in pixels or centimetres.
PCK / PCKh	Mainly cattle, then pigs and horses	Common in temporal pose estimation, markerless 2D pose regression, and anatomical landmark localisation where normalisation by head/body size is meaningful.
Accuracy / F1 / Sensitivity / Specificity / AUC / Kappa	Mainly cattle, then poultry, horses, and pigs	Most common in pose-informed downstream classification, including lameness, mobility scoring, behaviour recognition, and health screening tasks.
Recall / AR / mAR	Mainly cattle, then poultry and pigs	Used both in detection/ keypoint benchmarking (AR/ mAR) and in behaviour/ classification pipelines (Recall). Frequently paired with AP-family metrics.
Correlation / agreement metrics	Mainly cattle, with some horses and pigs	Typical of 3D pose, measurement, and clinical agreement studies, especially where automated outputs are compared with human scores, manual measurements, or reference systems.
Efficiency metrics (FPS / inference speed / latency)	Mixed, with examples in poultry, pigs, goats, camels, and cattle	Most often reported in lightweight top-down/direct models and deployment-oriented edge pipelines.
Model complexity metrics (Parameters / model size / FLOPs)	Mixed, with many examples in cattle, plus pigs, poultry, goats, horses, and camels	Most common in studies proposing lightweight architectures, real-time pose models, or edge-deployment variants.
Temporal robustness metrics (TC / VCP)	Mostly cattle in the supplied studies	Reported mainly in tracking-aware or structural post-processing pipelines assessing frame-to-frame stability and usable-output rate.

3.2.3. *Cost-benefit synthesis across sensing configurations*

Formal return-on-investment analyses are still rare in the reviewed literature. However, the available studies support a practical cost-benefit hierarchy. RGB single-camera systems offer the lowest hardware and installation burden and are best suited to large-scale monitoring of behaviour, mobility, feeding, or identity in well-chosen viewpoints. RGB-D and stereo systems occupy a middle ground: they increase hardware and calibration cost but offer clear gains for body measurement, weight estimation, and some lameness or posture tasks in which metric depth matters. Multi-view RGB or RGB-D rigs provide the strongest occlusion robustness and three-dimensional recoverability, but they also impose the highest installation, synchronisation, maintenance, and data-management costs. A list of the most used camera devices and their main characteristics is summarised in Table 9.

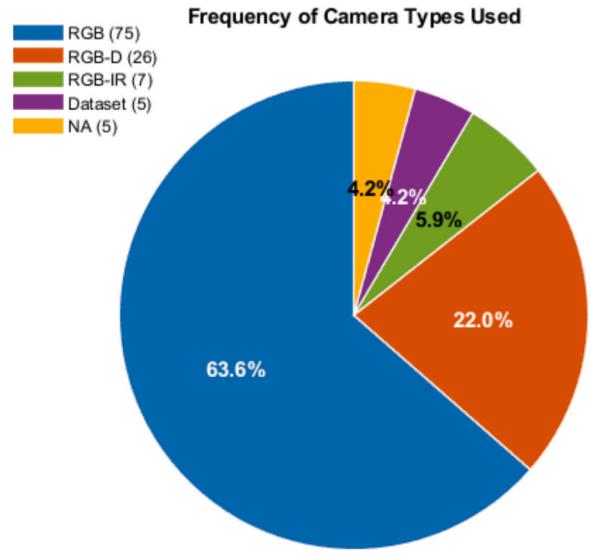

Fig. 12. Distribution of camera types across the included studies.

Table 8
Number of studies that use multiple cameras and multiple camera technologies.

No. of cameras	Type of camera	No of works
7	RGB	3
5	RGB; RGB-D	2
4	RGB; RGB-IR	3
3	RGB; RGB-D	6
>2*	RGB; RGB-IR	3
2	RGB; RGB-D; RGB-IR	10

* exact number not specified.

In practical terms, the literature suggests that large-area pen or corridor monitoring should default to fixed RGB when relative posture or event detection is the primary objective, whereas targeted high-value locations such as weighing corridors, grouping platforms, breeding stations, or dedicated welfare-monitoring zones can justify RGB-D, stereo, or multi-view deployments. This task-location matching strategy is better supported by the evidence than any "one-sensor-for-all-tasks" approach.

3.3. *How depth-enhanced RGB systems and multi-camera setups compare with RGB-only approaches (RQ.3)*

The revised corpus confirms that depth-enhanced and multi-camera systems are not universally superior, but they do provide systematic advantages under three recurring conditions: when absolute geometry is required, when occlusion is intrinsic to the scene, and when camera-animal distance or background variation would otherwise destabilise 2D inference.

For morphometrics and non-contact measurement, depth or stereo usually reduces ambiguity and improves reproducibility. Across cattle, pigs, sheep, horses, poultry, and yaks, depth-backed pipelines typically report low relative errors for body dimensions when posture can be constrained or filtered (Du et al., 2022; Luo et al., 2023; Li et al., 2024c; Peng et al., 2024; Yang et al., 2025; Deng et al., 2025; Yue et al., 2025; Zhao et al., 2025b; Wei et al., 2025; Li et al., 2025b; Xu et al., 2025a).

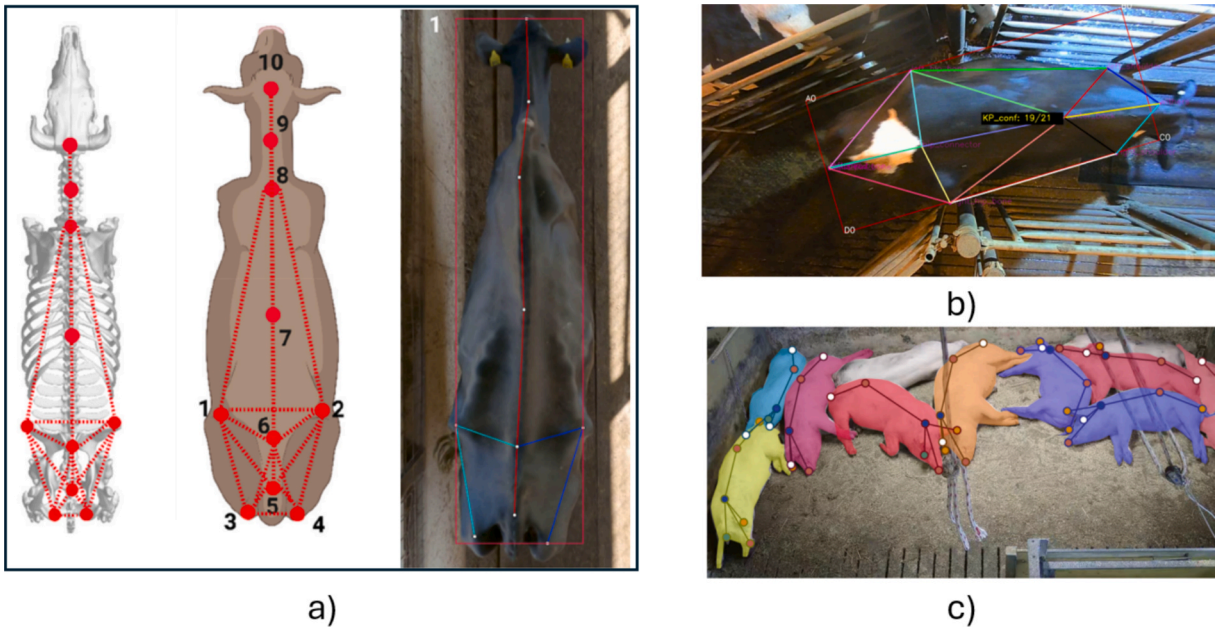

Fig. 13. Representative examples of top-down imaging configurations used for livestock keypoint detection and monitoring: (a) Menezes et al., 2025; (b) Ramesh et al., 2023; (c) Kresovic et al., 2022.

Even when pose is estimated in 2D, depth back-projection converts keypoints into physically meaningful measurements, which is difficult to achieve robustly from monocular RGB alone.

For occlusion-sensitive behaviour and posture analysis, multi-view systems are especially valuable. In dairy-cow freestall monitoring, synchronized multi-camera fusion enabled sub-second timing of lying and rising phases with strong agreement to human observers (Kroese et al., 2024b, 2025). In cattle monitoring at barn scale, multi-camera bird's-eye-view mapping improved global spatial consistency across overlapping and non-overlapping views (Nasir et al., 2025) (Fig. 14). In re-identification, viewpoint-aware filtering and pose-guided attention improved robustness across camera changes and posture variation (Liu et al., 2025b). These studies show that multi-view sensing is most defensible when the single-view alternative would fail for structural reasons rather than because of backbone weakness alone.

For RGB-only systems, the main advantages remain simplicity, cost, and ease of scaling. RGB can perform extremely well when the scene is controlled, when the target variable is primarily relative rather than metric, and when the keypoint schema is compatible with the visible anatomy. Side-view RGB is therefore often sufficient for mobility and lameness screening, while top-down RGB often provides the best trade-off for group-housed behaviour, activity, and identification. The strongest evidence does not suggest replacing RGB everywhere, but rather augmenting it selectively when the endpoint demands geometric certainty or the housing geometry creates unavoidable blind spots (Table 10).

3.4. Machine learning architectures and training strategies most effective for livestock pose-based monitoring (RQ.4)

The updated literature supports a predominantly modular design philosophy (Table 11). In most successful farm-oriented systems, keypoint localisation is separated from downstream inference. Keypoints are first detected using a dedicated pose model and are then converted into behaviour classes, gait descriptors, morphometric measurements, or identity features through classifiers, regressors, or geometric rules. This design is especially common in cattle mobility scoring, lameness analysis, and body measurement because it preserves clinical interpretability and allows the pose front-end to be replaced without redesigning the downstream task logic (Zhao et al., 2023a; Higaki et al., 2024; Russello et al., 2024; Narli et al., 2025; Reitmaier and Narli, 2025; Zhao et al., 2025a).

For keypoint localisation itself, DeepLabCut, HRNet, YOLOv8-Pose/YOLOv11-Pose, RTMPose, and related derivatives remain the most influential families. DeepLabCut remains particularly common in animal-specific, small- to medium-scale academic datasets because of its transfer-learning accessibility and strong single-instance performance. HRNet and related high-resolution architectures remain highly competitive for precise skeletal localisation, especially in poultry, pigs, and livestock measurement tasks. YOLO- and RTM-based systems have become increasingly important where real-time operation, simpler deployment, or tighter integration with detection are needed.

Training strategies are still dominated by supervised learning with transfer learning from ImageNet, COCO, AP-10 K, Animal-Pose, or related pretrained backbones (Table 12). However, the updated 2025 literature places greater emphasis on transferability and domain adaptation. MT-SRNet explicitly uses active learning and scene adaptation across pig farms (Liu et al., 2025a), SURABHI improves cattle recognition by rectifying hard instances in a self-training loop (Ramesh and Reibman, 2024), and several recent works explicitly identify synthetic data, pseudo-labeling, domain alignment, or self-supervised learning as the next step for handling cross-farm shift (Liu et al., 2025a; Wu et al., 2025; Reitmaier and Narli, 2025; Li et al., 2025a). This indicates that cross-site robustness is becoming a central concern rather than a secondary one.

Task-specific feature engineering remains especially effective when the phenotype is well understood biomechanically. Lameness systems repeatedly benefit from explicit gait descriptors such as back curvature, stride timing, head swing, limb asymmetry, or dorsal sway rather than end-to-end black-box prediction alone (Zhao et al., 2023a; Higaki et al., 2024; Russello et al., 2024; Narli et al., 2025; Reitmaier and Narli, 2025; Duan et al., 2025; Wu et al., 2025). By contrast, interaction recognition, multi-task posture segmentation, or re-identification benefit more from representation learning, graph modelling, or transformer-style feature fusion (Yang et al., 2023; Liu et al., 2025b).

Transfer learning is beneficial in livestock pose estimation mainly because most target datasets remain small, species-specific, and expensive to annotate. Pre-training on ImageNet or COCO can improve

Table 9
List of the most used (>3 works) camera systems across the examined works.

Camera	No. of Works	Type of camera	Characteristics
Hikvision/ Ubiquiti surveillance camera	29	RGB/ RGB-IR	Network/IP security cameras (often with IR night mode) CCTV cams. Pros: 24/7, weather-rated, long runs over Ethernet. Cons: heavy compression, rolling shutter/auto-exposure quirks, mixed optics; SDK limitations may restrict system integration.
Intel RealSense (D415-D435i)	11	RGB-D	Active stereo IR depth + RGB. Pros: decent depth indoors, good SDK, workable on edge devices. Cons: IR depth struggles in sunlight/steam/dust; specular coats can cause holes.
Microsoft Kinect (V2/azure)	7	RGB-D	Active time-of-flight depth + RGB. Pros: accurate depth for gait/ posture, good SDKs. Cons: indoor-biased, sensitive to sunlight/ reflective coats, discontinued v2 hardware.
Stereolabs ZED (2-2i)	6	RGB-D	Passive stereo RGB that computes depth from disparity. Pros: long-range depth outdoors (no IR projector to wash out), wide FOV. Cons: needs texture/lighting; depth degrades on uniform coats/ motion blur.
GoPro (Hero6-9)	6	RGB	Action cams, high-FPS 4 K (and up), ultra-wide FOV, strong stabilisation. Pros: low-cost, portable, suitable for mobile data capture and high-frame-rate kinematics. Cons: rolling shutter, heavy lens distortion (needs calibration), low-light noise; no native depth.
Apple iPhone (8–12)	4	RGB	High-quality RGB video; strong ISP and stabilisation. Pros: easy, ubiquitous, high-quality colour reproduction and exposure control; useful for field data collection. Cons: variable control over auto settings.

convergence and low-level visual feature extraction, particularly for edges, textures, body contours, and generic object localisation. However, such transfer should be interpreted primarily as visual-feature reuse rather than anatomical transfer. Human-pose datasets encode upright bipedal topology, limb proportions, joint visibility patterns, and camera viewpoints that differ substantially from quadrupeds and birds. As a result, direct transfer from human-pose sources can reduce annotation burden but may also introduce anatomical bias, particularly for dorsal views, distal limb landmarks, occluded joints, tails, ears, or species-specific body regions that have no human equivalent.

Animal-pose pre-training sources such as AP-10 K or Animal-Pose provide a closer morphological prior than human-pose datasets, but they are still not automatically aligned with production livestock conditions. Many generic animal datasets contain companion animals, wildlife, or zoo-like imagery rather than barn, pen, corridor, pasture, or cage environments. Therefore, their usefulness depends on the similarity between source and target species, viewpoint, keypoint taxonomy, and background conditions. Cross-livestock transfer is most defensible between morphologically or visually related domains, such as cattle and yak, adult pigs and piglets, or similar top-down pen-monitoring tasks. In contrast, transfer from cattle to poultry, from side-view gait analysis to top-down identity monitoring, or from clean laboratory images to dusty commercial housing may lead to negative transfer unless domain adaptation, fine-tuning, or target-domain validation is performed.

Overall, the evidence suggests that transfer learning should not be treated as a plug-and-play solution. Future livestock pose-estimation studies should report the pre-training source, the mapped keypoints, the number of target-domain annotations used for fine-tuning, whether layers were frozen or fully updated, and performance on unseen farms or breeds. This would make it possible to distinguish genuine cross-species generalisation from successful fine-tuning on narrowly matched target data.

3.5. *Data bottlenecks, annotation workflows, and dataset solutions (RQ.5)*

The updated corpus confirms that data scarcity is not simply a matter of dataset size, but of diversity, annotation burden, and inconsistent taxonomies. Many studies remain single-site, single-breed, or single-view, which limits transfer to new farms or housing types. Even when dataset sizes are large, they are often private or partially released, and the keypoint schema may be too task-specific for wider reuse. This remains one of the main reasons why strong results frequently degrade in unseen environments.

At the same time, the new 2025 studies show that dataset maturity is

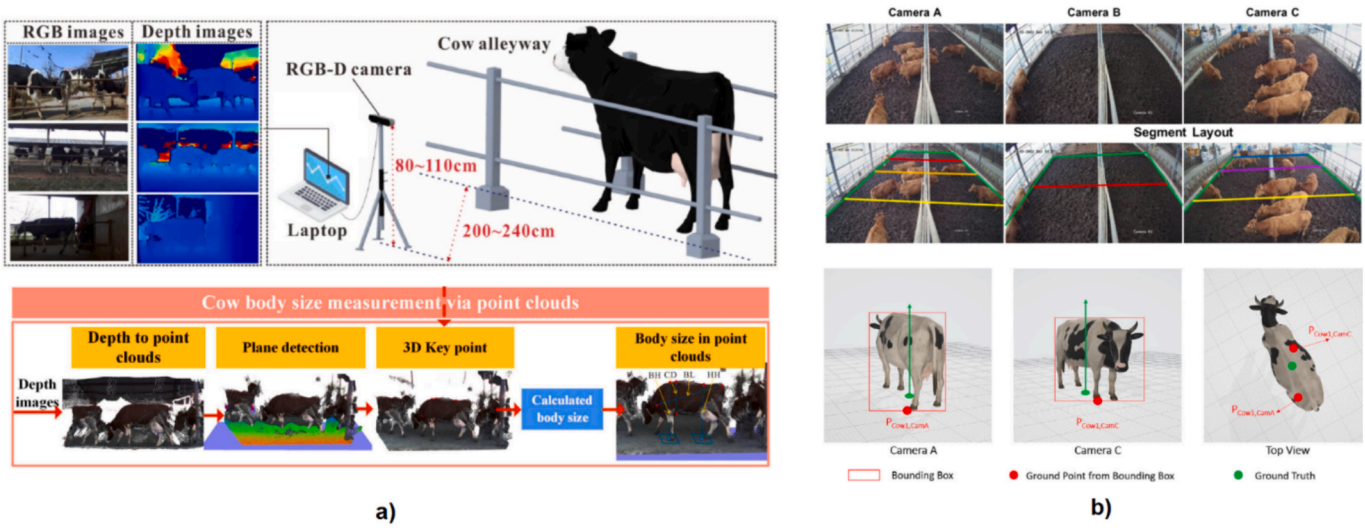

Fig. 14. A) rgb-d images were collected from operational dairy farms. for images used for measuring body size, barriers were removed to avoid covering the keypoints (Yang et al., 2025); b) Multi-camera approach, with 3D reconstruction of the cow body and its reaction force (Nasir et al., 2025).

Table 10
Summary of the advantages and disadvantages of the different imaging technologies.

System Type	Strengths	Limitations	Best Use Cases
RGB-only	Low-cost, pre-trained model support, low complexity	Sensitive to occlusion and lighting; depth ambiguity	Controlled environments, early prototyping
RGB-D	Better spatial awareness, improved occlusion handling	Infrared interference in sunlight, higher power needs	Indoor barns, crowded pens
Multi-camera	Resolves occlusion, allows 3D triangulation	Calibration and sync challenges; high data load	Complex behaviour tracking, group analysis
Hybrid	Multi-modal resilience, temporal continuity	Integration complexity	Real-time systems in varied environmental conditions

Table 11
List of most used ML/DL model families (pose/detection/tracking).

Model family	Typical strengths / why widely used	Representative works
YOLOv8 / YOLOv8-Pose *(detection + pose)*	Fast one-stage inference and easy deployment.	Yue et al., 2025; Peng et al., 2024; Yang et al., 2024; Witte et al., 2024; Ong et al., 2023; Nasir et al., 2025; He et al., 2025; Zhao et al., 2024
DeepLabCut (ResNet backbones) *(pose)*	Transferable across species and effective with modest annotation effort.	Liu et al., 2020; Luo et al., 2023; Russello et al., 2022; Farahnakian et al., 2021, 2024; Niknejad et al., 2023; Khin et al., 2022; Du et al., 2022; Gorssen et al., 2022
HRNet / HRNet-based models *(pose)*	Strong spatial precision for keypoint localisation.	Kroese et al., 2025; Martin-Cirera et al., 2024; Wang et al., 2023b; Zhao et al., 2025a; Högberg et al., 2025
Mask R-CNN / Keypoint R-CNN *(detection + keypoints / segmentation)*	Mature two-stage framework with strong instance separation.	Barney et al., 2023; Ramesh et al., 2023; Ramesh and Reibman, 2024; Liu et al., 2023b
Stacked Hourglass / OpenPose-style frameworks *(pose)*	Established multi-stage heatmap pipelines, still robust in farm scenes.	Li et al., 2019; Yik et al., 2020; Kresovic et al., 2022; Menezes et al., 2025
RTMPose / YOLOv11-Pose / MobileViT-Pose *(lightweight pose)*	Designed for fast, compact, edge-oriented pose estimation.	Su et al., 2024; Li et al., 2023; Chen et al., 2025a; Deng et al., 2025; Xu et al., 2025a; Wu et al., 2025
PointNet++ / 3D GCN-based pipelines *(point-cloud pose / measurement)*	Suitable for direct 3D landmarking and morphometrics.	Xu et al., 2025b; Lu et al., 2025; Okour et al., 2024
Multiview / stereo 3D fusion pipelines *(3D pose / measurement)*	Improve depth recovery and geometric robustness.	Kroese et al., 2024b, 2025; Du et al., 2022; Deng et al., 2025; Wei et al., 2025

Table 12
List of most used pretrained sources / backbones.

Pretrained source	Typically paired with	Representative works (examples)
ImageNet *(ResNet-50/101, EfficientNet/-V2, MobileNet)*	Common backbone initialisation for pose and classification models.	Liu et al., 2020; Zhao et al., 2023a; Farahnakian et al., 2024; Siachos et al., 2024; Niknejad et al., 2023; Yang et al., 2025
COCO *(detection / pose)*	Standard transfer source for detection and keypoint pipelines.	Barney et al., 2023; Russello et al., 2024; Ong et al., 2023; Ramesh et al., 2023
HRNet pretrained weights *(COCO / AP-10 K)*	Widely reused for animal pose fine-tuning.	Kroese et al., 2025; Högberg et al., 2025; Martin-Cirera et al., 2024; Wang et al., 2023b
Ultralytics YOLO family *(v5 / v8)*	Practical starting point for fast farm-specific detection and pose models.	Nasir et al., 2025; Yue et al., 2025; Peng et al., 2024; Yang et al., 2024; Witte et al., 2024
ETH3D / stereo-depth weights	Used for stereo depth estimation modules.	Deng et al., 2025
MPII Human Pose / AP-10 K	Useful for legacy human-pose transfer and broad animal-pose pretraining.	Li et al., 2019; Ong et al., 2023
None / trained from scratch	Chosen for highly task-specific or custom architectures.	Perneel et al., 2025a; Weng et al., 2023

improving. Examples include PIG-ALL and related subsets for multi-farm pig monitoring (Liu et al., 2025a), HanwooReID for multi-view cattle re-identification (Liu et al., 2025b), YakPoseData for complex yak-heifer scenes (Li et al., 2025a), BECA for long-term dorsal-view beef-cattle recognition and pose (Zhang et al., 2025), and task-specific public or semi-public releases for goat aggression, pig feeding, or poultry monitoring (Qu et al., 2025; Hu et al., 2025; Xu et al., 2025a). These additions show a gradual move toward broader datasets, but the overall landscape is still fragmented by species, viewpoint, and task.

Annotation cost remains high, especially for dense multi-animal scenes and anatomically fine-grained taxonomies. Manual annotation is still dominant, but several practical strategies recur: frame selection by k-means or uncertainty sampling, active-learning loops, semi-automatic label correction, visibility flags for partially occluded joints, rule-based refinement, and synthetic or transferred annotation pipelines. Examples include active-learning-driven pig annotation in MT-SRNet (Liu et al., 2025a), semi-automated RGB-to-thermal annotation transfer in CattleFever (Pham et al., 2025), DIAT transfer from infrared to depth images in pigs (Yik et al., 2020), and rule-based or shape-prior refinement for cattle keypoints in SURABHI (Ramesh and Reibman, 2024). These approaches do not remove the need for expert annotation, but they clearly reduce manual effort.

A practical comparison of annotation ecosystems also emerges from the corpus. CVAT and LabelMe are the most common general-purpose tools because they support flexible keypoint and box annotation and fit COCO-like pipelines. Roboflow is attractive for browser-based annotation and augmentation, particularly in newer YOLO workflows. DeepLabCut's internal interfaces remain common in animal-behaviour studies using DLC-centred pipelines. CloudCompare is particularly useful for point-cloud-based annotation in 3D measurement studies. In practical terms, CVAT and LabelMe are well suited to medium-to-large custom pose datasets, Roboflow is convenient for rapid iteration and YOLO-based workflows, DeepLabCut interfaces are efficient for DLC-specific research pipelines, and CloudCompare is preferable when the primary annotation object is a 3D point cloud rather than a 2D image.

To make the discussion of annotation ecosystems more practical, Table 13 compares the main annotation tools and workflows identified across the reviewed studies, highlighting their strengths, limitations, and most suitable use cases. Rather than listing software only, the table is intended as a decision-support aid for selecting annotation strategies according to dataset size, data modality, and task complexity.

Overall, the reviewed studies indicate that lightweight tools such as LabelMe remain attractive for small and medium-sized 2D landmark datasets because of their flexibility and low setup burden, whereas CVAT is more suitable for larger collaborative projects, especially when video interpolation, identity labels, and visibility flags are needed.

Table 13
Comparison of annotation tools and workflows used in livestock pose-estimation studies: capabilities, limitations, and recommended use cases.

Tool / workflow	Typical annotation target	Main strengths	Main limitations	Best-suited scenario	Representative examples from this review
LabelMe	Keypoints, bounding boxes, simple masks	Lightweight, flexible, easy for manual keypoint placement, low entry barrier	Limited collaboration, automation, and interpolation; less efficient for large video-heavy datasets	General 2D livestock keypoint datasets; single-user or lab-scale annotation	Chen et al., 2025c; Gan et al., 2021; Li et al., 2024c, 2024b; Yang et al., 2024; Peng et al., 2024
CVAT	Keypoints, boxes, IDs, visibility flags, tracking labels	Strong video support, interpolation tools, robust export, suitable for team workflows	Heavier setup and management overhead than lightweight tools	Multi-user projects; video-based datasets; tasks requiring IDs, visibility flags, or tracking	Taghavi et al., 2024; Nasir et al., 2025; Qu et al., 2025; Chen et al., 2024; Liu et al., 2025b
DeepLabCut / DLC GUI	Markerless anatomical landmarks for pose estimation	Directly integrated with DLC training and refinement; efficient for iterative relabelling	Less suitable for general annotation tasks, boxes, or team-based workflows	Landmark-centred studies built directly around DeepLabCut training and retraining	Du et al., 2022; Wu et al., 2024; Russello et al., 2022; Klecel et al., 2025; Shirey et al., 2023; Feuser et al., 2022
Roboflow	Boxes, keypoints, augmentation-oriented datasets	Convenient browser-based workflow, dataset management, and export pipeline	Less specialised than CVAT for complex long-sequence annotation	Rapid dataset creation and augmentation for 2D pose tasks	Qu et al., 2025; Xu et al., 2025a
COCO Annotator	COCO-style keypoints and boxes	Good for standardised COCO-format annotation and model interoperability	Less feature-rich for long videos, QA, and workflow control than CVAT	Studies prioritising COCO compatibility and straightforward pose-model training	Hu et al., 2025; Martin-Cirera et al., 2024; Li et al., 2023; Ramesh et al., 2023
CloudCompare	3D keypoints, region labels, segmented anatomical structures	Well suited to point-cloud inspection, 3D landmarking, and morphometrics	Not designed for standard 2D image workflows; steeper learning curve for non-3D users	3D livestock body measurement, point-cloud segmentation, and geometric landmarking	Xu et al., 2025b; Wei et al., 2025; Li et al., 2025b
Label Studio	Keypoints, boxes, mixed labels	Flexible schema design and practical web interface for mixed annotation tasks	Less commonly used than LabelMe or CVAT in this literature	Mixed-label projects combining object and landmark annotation	Lodkaew et al., 2023; Yu et al., 2024
VGG Image Annotator (VIA)	Boxes, keypoints, identity labels	Simple, lightweight, browser-based, minimal setup	Limited automation and weaker support for large collaborative or video-heavy workflows	Structured box/keypoint annotation with minimal setup	Menezes et al., 2025; Zhao et al., 2025b
Supervisely	Boxes, keypoints, segmentation labels	Strong visual interface and practical export for linked pose-segmentation tasks	Less frequently used in this review; may be unnecessary for simple keypoint-only studies	Pose studies linked to segmentation or injury localisation	Volkmann et al., 2022
VoTT	Boxes, keypoints, IDs	Useful for structured frame annotation and identity-aware tracking workflows	Less suited to dense anatomical keypoint annotation than CVAT	Tracking benchmarks and identity-aware frame annotation	Bibinbe et al., 2025
Sloth	Bounding boxes	Simple manual bounding-box annotation	Limited functionality compared with newer collaborative tools; not pose-specialised	Basic detection datasets or preliminary object-labelling stages	Riekert et al., 2020
X-AnyLabeling	Tracking IDs, detection/ keypoint support	Useful complementary tool for ID annotation and post-labelling refinement	More niche and less standardised than CVAT or LabelMe	Tracking-focused post-processing or ID annotation	Chen et al., 2025a
Keylabs.ai	Boxes, masks, keypoints, visibility tags	Supports structured large-scale annotation and visibility tagging under QA workflows	Less established in academic livestock studies; platform-dependent workflow	Large curated datasets with boxes, masks, and keypoints under controlled QA	Ong et al., 2023
Custom tools / custom GUIs	Task-specific keypoints, clinical landmarks, bespoke IDs or structural templates	Highly adaptable to anatomy-specific problems and study-specific rules	Harder to reproduce, reuse, maintain, and standardise across studies	Bespoke anatomical schemas, clinical validation studies, or non-standard pose problems	Nasiri et al., 2022; Li et al., 2019; Lawin et al., 2023; Li et al., 2019; Hewitt and Mahmoud, 2019

DeepLabCut's built-in annotation interface is efficient for closed-loop landmark refinement within markerless pose workflows, but is less general-purpose than dedicated annotation platforms. For 3D morphometric studies, CloudCompare remains the most appropriate specialised option. More recent workflows based on active learning, model-assisted correction, and semi-automatic refinement appear particularly promising for large-scale livestock datasets, where manual annotation cost increasingly becomes a bottleneck.

Overall, progress in RQ.5 is increasingly driven by workflow design rather than annotation alone. The most promising direction is not simply larger datasets, but datasets that are more diverse across housing, lighting, breed, and viewpoint, combined with active learning, synthetic augmentation, and explicit taxonomy standardisation.

3.6. Explainability and interpretability in livestock pose estimation (RQ.6)

Most livestock pose systems still achieve interpretability through structural transparency rather than through complex post-hoc explainability methods. In practical terms, users are commonly shown skeleton overlays, keypoint trajectories, geometric measurements, or feature rankings rather than saliency maps alone. This is often appropriate because many livestock tasks are already naturally interpretable at the level of anatomy and motion.

Three broad interpretability modes recur. First, tracking evidence: skeleton overlays, confidence maps, temporal traces, and keypoint trajectories allow direct verification that the relevant anatomy has been detected and followed correctly. Second, feature-based evidence: when random forests, SVMs, boosting methods, or geometric rules are used,

the resulting feature importance or threshold logic often maps directly to biologically meaningful motion patterns. Third, geometric evidence: in body measurement and weight-estimation systems, the outputs are often explicit distances, widths, heights, volumes, or contour-derived circumferences in physical units, which makes the inference process inherently more transparent than an opaque end-to-end score.

The 2025 additions reinforce this pattern. Narli et al. (2025) and Reitmaier and Narli (2025) use engineered back-shape features and permutation or hierarchical analysis to identify the most informative spinal regions for lameness. Qu et al. (2025) explicitly defines aggression through head speed, direction, and inter-animal distance rather than latent sequence embeddings alone. Högberg et al. (2025) uses keypoint-to-brush geometry and spline projection to infer brushed body segments. Wang et al. (2025), Zhao et al. (2025b), Li et al. (2025a), and Xu et al. (2025a) all convert explicit keypoints into physical measurements, preserving a transparent link between detected anatomy and the final trait estimate.

True post-hoc explainable artificial intelligence (XAI) remains limited. Some studies use attention maps, CAM-like outputs, or ablation studies, but explicit user-centred XAI remains the exception rather than the rule. The literature therefore suggests that the near-term priority is not necessarily more sophisticated XAI, but clearer reporting of keypoint definitions, visibility handling, feature construction, and failure modes, all of which are more directly relevant to farmer or veterinarian trust.

3.7. Environmental and species-specific factors affecting accuracy and generalisability (RQ.7)

Species morphology and housing context strongly condition what pose estimation can realistically achieve. Cattle studies dominate partly because many economically important tasks, such as lameness screening, body measurement, and identification, align well with corridor or parlor viewpoints. Pigs dominate pen-scale top-down studies because group housing, feeding, and welfare monitoring are naturally suited to overhead observation. Poultry studies often struggle with crowd density, rapid motion, feather uniformity, and small distal anatomy. Horse studies benefit from larger body size and more controlled clinical or breeding settings, but still face difficulties with stride variability, outdoor lighting, and single-view occlusion. Goat, sheep, yak, camel, and turkey studies remain fewer, and often still task-specific.

Environmental factors recur with striking consistency. Occlusion remains the dominant challenge in multi-animal housing. Lighting variation, heat lamps, reflective metal structures, cage nets, dust, and background clutter repeatedly destabilise RGB-only systems. Long hair, homogeneous coat colour, and breed-dependent body proportions can reduce keypoint visibility or increase localisation uncertainty. Several recent studies show this explicitly: yak hip and leg keypoints remain difficult because of long hair (Li et al., 2025a); duck and poultry systems degrade under occlusion and rapid motion (Zhao et al., 2023b, 2025a); high-angle dairy-cow brush monitoring performs substantially worse than side or top-down views (Högberg et al., 2025); and side-view weight or measurement systems remain sensitive to angle variation and background quality (Rozendo et al., 2025; Li et al., 2024c).

Generalisation is strongest when the camera geometry and keypoint taxonomy are matched to the species and task, rather than transferred naively. Systems trained on top-down pig or cattle views do not automatically transfer to side views or to dorsal-only identity tasks. Likewise, schemas optimised for lameness are often unnecessarily dense for feeding or occupancy analysis, whereas compact dorsal schemas used for identity or mobility would be insufficient for detailed distal-joint gait analysis. The strongest evidence therefore supports task-specific rather than “generic” keypoint design.

3.8. Main deployment challenges in real farm environments (RQ.8)

The main barriers to deployment are operational rather than purely algorithmic. The most frequent failure points across studies are camera geometry, occlusion, animal density, calibration drift, environmental survivability, and limited edge compute, rather than the nominal choice between one deep backbone and another. This is consistent across cattle, pig, poultry, goat, and horse applications.

Representative real-farm case studies show both the promise and the constraints of deployment. Zhao et al. (2025b) integrated RGB-D keypoint detection, RFID, dynamic weighing, and automated sorting into a beef-cattle grouping platform, demonstrating how vision can be linked directly to management action. Li et al. (2025b) built a portable RGB-D horse measurement device with battery-powered Jetson processing and optional manual correction, showing a credible route to low-stress field deployment beyond laboratory conditions. Högberg et al. (2025) demonstrated welfare-oriented brush-use monitoring with commodity cameras in loose housing, but also showed how error propagation from one stage can still constrain full automation. Wang et al. (2025) demonstrated low-cost edge-assisted yak weight estimation in plateau conditions, but also exposed the limits of Raspberry Pi inference speed for richer video analytics.

Edge hardware evidence is now sufficient to support a comparative synthesis. Jetson-class devices can already support several practical pose pipelines at or near real-time for selected tasks, as shown by goat aggression monitoring, horse RGB-D measurement modules, and cattle body-measurement systems (Qu et al., 2025; Li et al., 2025b; Weng et al., 2025). Raspberry Pi-class deployment remains feasible for lighter or snapshot-style inference, but is often too slow for complex continuous video pipelines without substantial pruning, quantisation, or temporal subsampling (Tan et al., 2025; Wang et al., 2025). These findings also support a practical hardware-selection guideline: lightweight 2D pose and event detection are increasingly realistic on embedded GPUs, whereas multi-view reconstruction, point-cloud processing, and richer regressors remain less practical on very low-power devices.

Integration with farm-management infrastructure is still under-reported, but the evidence that does exist suggests a pattern. Most deployed or near-deployed systems interface through RFID, local storage/network video recorder pipelines, or edge-to-dashboard workflows rather than through deeply standardised farm middleware. This is visible in grouping platforms, handheld or vehicle-based weight systems, breeding measurement stations, and long-term dataset capture infrastructures (Zhao et al., 2025b; Wang et al., 2025; Xu et al., 2025a; Zhang et al., 2025). The gap is therefore not whether integration is possible, but that communication protocols and interface design are rarely described in enough detail for straightforward replication.

3.9. Keypoint definitions across livestock species and implications of anatomy (RQ.9)

Keypoint definition is one of the most under-standardised aspects of the field. The updated corpus confirms large variation not only across species but within species and tasks. Cattle studies range from very compact dorsal or back-only schemas for mobility, identity, or lameness to dense limb-centric schemas for gait and morphometrics. Pig studies span minimal 2–6 point tracking schemas, 8–10 point body-measurement schemas, and denser 13–20 point skeletons for behaviour or action recognition. Poultry schemas vary from head-only or part-based detection to 8–10 point whole-body layouts. Horse studies range from compact clinical keypoint sets for lameness to richer 16- to 58-point anatomical schemes. Goat, sheep, yak, camel, and turkey studies remain few but already show notable variation in task-specific landmark definitions.

This heterogeneity matters because it directly shapes annotation effort, model complexity, metric behaviour, and biological interpretability. Distal-joint-rich schemas are useful for gait and lameness but are harder to annotate and more sensitive to blur and occlusion. Compact dorsal schemas are easier to detect and may be sufficient for mobility scoring, identity, occupancy, or brush-use analytics, but they cannot

recover detailed limb kinematics. Measurement-oriented schemas often use a hybrid logic in which a limited number of stable anchor keypoints are detected and the final measurement points are derived through contour or depth processing rather than predicted directly. This pattern is especially common in cattle, horse, poultry, and yak measurement systems.

A preliminary standardisation guide therefore emerges from the synthesis. For cattle, a minimal dorsal set should include head or poll, withers, thoracolumbar/back reference, hip or hook landmarks, and tail head for top-down mobility or identity tasks, while gait applications additionally require distal limb joints and hoof landmarks. For pigs, a practical minimum set for top-down behaviour and monitoring should include nose or snout, ears, neck or shoulder, hip or mid-back, and tail base, with limb additions where gait or detailed action recognition is required. For poultry, at least head or beak, trunk centre, tail, and leg/foot landmarks are needed for behaviour and gait, with denser schemas justified only when clinically relevant distal cues are required. For horses and other larger quadrupeds, head, withers, pelvis or sacral region, and distal limb landmarks remain the minimal clinically relevant set for locomotor analysis, while measurement applications benefit from added contour anchors.

This preliminary guide does not replace formal species standards, but it provides a concrete starting point for improving comparability, data sharing, and model reuse. It also supports the conclusion that anatomical standardisation is now a limiting factor at least as important as raw model capacity.

4. Discussion

4.1. Main advances and positioning of the review

The present review expands the evidence base on vision-based livestock pose estimation to 114 included studies and shows that the field has moved beyond proof-of-concept landmark detection toward application-oriented systems for mobility scoring, welfare analytics, morphometrics, weight estimation, behaviour recognition, tracking, and re-identification. The strongest general conclusion is that performance depends less on the nominal choice of backbone and more on the alignment among five design decisions: the biological question, the camera geometry, the keypoint schema, the temporal or geometric reasoning layer, and the operational constraints of the farm environment. Across the reviewed literature, systems perform best when these elements are co-designed rather than selected independently.

This conclusion extends prior reviews in three ways. First, it confirms that pose estimation should be treated as a distinct methodological layer within precision livestock farming rather than as a minor component of general behaviour recognition. Earlier reviews largely centred on behaviour classification, body measurement, or general animal pose literature, but did not fully resolve the design trade-offs specific to livestock pose pipelines under commercial conditions. Second, the updated synthesis shows more clearly than earlier reviews that view selection is not a secondary implementation detail: top-down RGB, side-view RGB, and depth- or multi-view systems are not interchangeable, because they expose different anatomical evidence and support different endpoints. Third, the present synthesis indicates that practical deployment is increasingly limited by data diversity, annotation consistency, and operational integration rather than by the absence of high-capacity neural architectures.

A useful way to interpret the evidence is to distinguish between three maturity levels of pose-based livestock monitoring. The first includes constrained tasks with stable viewpoints and limited ambiguity, such as top-down posture recognition in pigs, corridor-based lameness screening in cattle, or controlled body measurement at dedicated checkpoints. In these settings, current methods are already technically mature enough for practical deployment. The second includes semi-structured farm tasks in which environmental variation, moderate occlusion, and heterogeneous behaviour still challenge generalisation, such as brush-use analysis, loose-housing interaction monitoring, and group-pen feeding analytics. These applications are feasible but remain sensitive to camera placement, annotation design, and temporal error propagation. The third includes open, highly variable, or densely crowded scenarios in which posture, identity continuity, and geometry must all be recovered simultaneously under uncontrolled illumination, weather, or background conditions. For these settings, the evidence remains promising but not yet sufficiently mature for broad, low-maintenance deployment.

4.2. Task-camera alignment as the central design principle

One of the clearest findings is that camera geometry should be chosen by task rather than by convention. Side-view RGB remains the most defensible default for locomotion and lameness because sagittal asymmetries, stride timing, head bobbing, and back curvature are directly visible in that view. Top-down RGB is usually the best compromise for group-level activity, feeding, nursing, social cohesion, and identity-related applications because it reduces self-occlusion and scales well to larger monitored areas. Depth-enhanced RGB, stereo, or multi-view systems are most justified when the target trait is inherently metric, when occlusion is intrinsic to the housing geometry, or when the decision requires three-dimensional posture rather than relative movement alone. This pattern appears repeatedly across cattle, pigs, poultry, horses, goats, sheep, and yaks, and suggests that the most useful design principle for practitioners is not “use the most advanced sensor”, but “fit the sensor geometry to the biological endpoint.”.

The review also clarifies that top-down and bottom-up pose formulations should not be framed as rivals in the abstract. Top-down systems remain dominant because they are easier to train, more mature in available toolchains, and often highly accurate when animals can be separated at the detection stage. Bottom-up systems become more attractive when the monitored setting is intrinsically multi-animal, when the density of animals is high, or when herd-scale monitoring must continue despite overlap. In practical terms, the literature suggests that top-down systems remain appropriate for most checkpoint, corridor, and moderate-density pen applications, while bottom-up or association-heavy pipelines become increasingly important for dense group housing, long-duration tracking, and open multi-animal monitoring.

The updated evidence also supports a more nuanced view of temporal modelling. Sequence-aware methods are clearly valuable when the phenotype is inherently dynamic, such as lameness, gait change, posture transitions, aggressive interaction, or prolonged behavioural bouts. Temporal convolution, recurrent models, keypoint tracking, Kalman filtering, and graph-based temporal aggregation all improve robustness to partial visibility and reduce the instability of frame-wise errors. However, temporal modelling is not equally necessary for all tasks. For many morphometric, checkpoint, or single-frame classification problems, well-designed static pipelines remain sufficient. This matters for deployment because temporal models often increase compute load, annotation demands, and implementation complexity.

4.3. Standardisation, metrics, and annotation bottlenecks

A major obstacle identified throughout the review is the lack of standardisation in keypoint taxonomies, evaluation metrics, and annotation logic. The same species is often represented with very different landmark sets depending on whether the target is gait, feeding, body measurement, tracking, or identification. This diversity is understandable, but it also limits transfer learning, benchmark comparability, and dataset reuse. A preliminary standardisation path emerges from the reviewed studies. For cattle, a compact dorsal schema can support mobility, identification, and brush-use analytics, whereas detailed gait applications additionally require distal limb and hoof landmarks. For

pigs, a minimal top-down schema anchored on nose, ears, shoulder or mid-back, hip region, and tail base is often sufficient for feeding, posture, and group monitoring, whereas detailed gait or morphometric tasks need denser body or limb representations. Poultry applications typically require head, trunk, tail, and leg or foot landmarks as a minimum, while horses benefit from compact clinically meaningful locomotion anchors plus distal limb points for lameness. The key implication is that species-specific core sets should be defined around biological use cases rather than around inherited human-pose conventions.

Metric heterogeneity remains another major barrier. The review confirms that many studies still report only pose-localisation metrics or only downstream task metrics, which makes practical comparison difficult. A pose estimator with a strong AP or PCK score may still be unsuitable for clinical lameness screening if temporal consistency is poor, while a system with excellent weight-estimation error may be difficult to compare if the physical-unit error is reported without any pose-quality metric. The evidence therefore supports a task-aligned reporting standard. Behaviour and posture studies should pair pose accuracy with macro-F1 or per-class recall. Lameness and mobility studies should report pose localisation together with balanced accuracy or macro-F1, sensitivity, specificity, and a temporal consistency indicator. Morphometric and weight-estimation studies should pair keypoint localisation with MAE, RMSE, and MAPE in physical units. Tracking and re-identification studies should combine pose performance with IDF1, HOTA, MOTA, or Rank-1 and mAP as appropriate. Such a reporting structure would substantially improve comparability without forcing all studies into a single unsuitable metric.

The annotation literature suggests that the field is beginning to mature operationally, but is still constrained by high labelling cost and by the scarcity of multi-site open datasets. Most studies still depend on manual annotation, often at substantial expert cost, especially when distal joints, thermal facial landmarks, or dense multi-animal scenes are involved. Nevertheless, several important methodological trends are visible: active learning to prioritise difficult frames, machine-assisted correction loops, visibility-flag conventions, annotation transfer across modalities, self-training with rectified hard cases, and prompt-guided or mask-guided scene adaptation. These workflows suggest that the future bottleneck is less likely to be raw model capacity than the efficiency and reproducibility of dataset construction. Public release of data, annotation schemas, and code remains inconsistent, and this inconsistency is now one of the main barriers to external validation.

4.4. Deployment readiness and integration into farm workflows

High pose-estimation accuracy in curated datasets does not necessarily imply readiness for commercial farm deployment. Farm environments introduce variable illumination, dust, dirt, steam, occlusion, changing backgrounds, inconsistent animal orientation, camera obstruction, and practical constraints related to power, connectivity, calibration, cleaning, data storage and operator interaction. Deployment readiness should therefore be assessed not only through keypoint-localisation or classification metrics, but also through the reliability of the complete sensing, processing and decision-support pipeline.

The reviewed literature suggests two main deployment frameworks. The first is a checkpoint phenotyping framework, suitable for body measurement, weight estimation, body condition assessment, lameness screening or sorting decisions in cattle, horses, yaks and other medium-to-large animals. In this case, cameras are placed at controlled locations such as races, corridors, weighing stations, parlour exits or feeding gates. The system may combine RGB, RGB-D, stereo or side-view cameras with RFID or visual identification, dynamic weighing, local or edge processing, and a fixed or portable user interface. The data flow begins when an animal enters the measurement zone: the animal is identified, frames are filtered for acceptable posture and visibility, keypoints or 3D anchors are extracted, and the system outputs body dimensions, weight-related traits, locomotion indicators or health flags. These outputs can support nutritional adjustment, sorting, veterinary inspection, breeding decisions or follow-up measurement. The main operational problems are frame-selection reliability, calibration drift, missing or noisy depth data, dust, throughput limitations, identity mismatches and the need for manual correction or operator override.

The second is a continuous welfare-monitoring framework, suitable for group-housed pigs, poultry, goats, dairy cattle and loose-housing systems. Here, fixed top-down, oblique or side-view cameras continuously monitor pens, cages, barns, feeding areas or welfare-related zones. The system typically includes animal detection and tracking, pose estimation, temporal aggregation, event detection, dashboard visualisation and alert logic. Pose trajectories are converted into higher-level indicators such as lying, standing, feeding, drinking, walking, aggression, social interaction, brush use, abnormal inactivity or changes in locomotor pattern. These outputs can help farmers prioritise manual checks, confirm welfare events, detect abnormal behaviour or monitor recovery. The main operational challenges are animal overlap, identity switches, occlusion, lighting variation, camera obstruction, dust accumulation, night-time imaging and gradual domain shift over time.

In both frameworks, user feedback is essential. Farmers, veterinarians or technicians may need to confirm alerts, correct identity errors, review rejected frames, adjust thresholds or trigger recalibration. A farm-ready system should therefore report not only pose accuracy, but also latency, hardware platform, calibration requirements, missing-data rates, tracking continuity, alert reliability, maintenance needs and integration with farm-management decisions. This shows that livestock pose estimation should be treated as one component of a wider decision-support system, rather than as an isolated computer-vision task.

4.5. Emerging directions, ethics, and governance

Emerging methodological directions deserve brief discussion because they are likely to shape the next generation of livestock pose systems. Vision Transformers are already visible in some recent work and appear especially promising for multi-scale reasoning, viewpoint robustness, and integration of pose with identity or context. Self-supervised and semi-supervised learning are also likely to become increasingly important because labelled livestock pose datasets remain expensive and narrow in scope. Synthetic-data pipelines and sim-to-real transfer are particularly attractive for 3D pose, occlusion-rich scenes, and rare species, although the reviewed evidence suggests that landmark-level validation remains essential because synthetic realism does not automatically guarantee anatomical fidelity. Diffusion-based methods are largely absent from the current livestock pose literature, but may become useful for data augmentation, occlusion completion, or generative regularisation. Federated learning is also scarcely represented in the reviewed corpus, yet it is a potentially important future direction because many farms or technology providers may be unwilling to centralise raw video data even when they would be willing to contribute model updates. These methods are therefore better framed as near-term opportunities than as current standards.

Ethics, privacy, and governance are still underdeveloped themes in the literature, despite their practical relevance. Unlike human surveillance, livestock pose systems do not usually raise direct personal privacy issues at the level of the animals themselves; however, they do raise important concerns around farm confidentiality, location inference, business-sensitive performance data, and access control to long-duration video streams. Some applications, such as identity-aware tracking, health monitoring, thermal sensing, or re-identification across sites, also increase the sensitivity of the generated data. The review therefore supports the inclusion of a basic governance layer in future deployment guidelines: minimisation of identifiable farm metadata in shared datasets, role-based access to raw and processed data, encryption of stored video and derived analytics where appropriate, and clear separation between public benchmarking assets and commercially sensitive farm records. In parallel, standardisation efforts should align with broader

precision-livestock-farming interoperability and FAIR (findable, accessible, interoperable, and reusable) data ambitions so that technical progress does not become trapped in isolated proprietary ecosystems.

The path to standard implementation is likely to require coordination across several actors rather than purely technical optimisation. A plausible promotion pathway begins with species- and task-specific benchmark datasets using harmonised landmark definitions and transparent reporting bundles, followed by multi-farm validation studies that report both technical metrics and operational key performance indicators. From there, deployment-oriented consortia involving researchers, breeders, veterinarians, farm-management software providers, and hardware vendors could support pragmatic reference workflows for common use cases such as corridor lameness screening, top-down feeding analysis, checkpoint morphometrics, or welfare-event detection. Without such a pathway, the field risks producing many strong point solutions that remain difficult to compare, integrate, or trust at scale.

4.6. Limitations of the current literature and of the present review

Several limitations of the present review should also be acknowledged. First, the literature remains highly heterogeneous in species, endpoints, camera configurations, and evaluation protocols, which limits the possibility of formal *meta*-analysis. Second, many datasets remain private or partially described, reducing reproducibility and making some claims difficult to verify externally. Third, some parts of the literature remain single-site and single-breed, so performance may be overestimated relative to wider deployment contexts. Fourth, although studies first indexed online in 2025 but assigned to 2026 issues were intentionally included to provide a complete 2015–2025 synthesis, this choice slightly increases the number of very recent studies with limited external replication. Finally, long-term field deployments, cost-effectiveness analyses, and cross-farm robustness studies remain scarce compared with the large number of technical development papers. These limitations do not weaken the central trends identified here, but they do mean that the precise boundaries of performance, cost, and generalisability remain provisional.

Overall, the literature now supports a more mature and practical view of livestock pose estimation (Table 14). The field is no longer defined only by whether keypoints can be detected, but by whether those keypoints are the right ones, captured from the right view, processed with the right temporal or geometric logic, and embedded into a workflow that is meaningful on farm. The strongest future progress is therefore likely to come not from ever more complex backbones alone, but from better task-specific standardisation, more diverse and openly described datasets, stronger deployment reporting, and tighter integration between pose analytics and real management decisions.

4.7. A decision framework for livestock pose-estimation system design

Based on the synthesis above, the proposed design framework can be interpreted as a decision process that converts farm and biological requirements into a concrete pose-estimation configuration (Fig. 15). The framework begins with the intended farm decision or phenotype, rather than with the model architecture. The first input is the task type: for example, lameness screening, feeding behaviour, social interaction,

Table 14
Practical design framework for livestock pose systems: from farm question to camera and model choice.

Target question / use case	Example traits / outputs	Recommended camera geometry	Recommended sensing modality	Pose strategy & model class	Practical notes
Herd-level mobility screening in corridors	3–4 level mobility/ lameness score, walking speed, lateral sway	Top-down or high-angle corridor view, 2.5–3.0 m height, 1080p, 20–30 fps	Commodity RGB; optional depth only if corridor geometry varies strongly	Top-down 2D pose (AiPEC/ YOLOv8-Pose/RTMPose-style), 10–25 dorsal keypoints; shallow classifier (RF/SVM) over gait features	Favours scalability and robustness to occlusion; works well at milking- or weigh-corridor chokepoints. Good default for farms wanting simple mobility analytics at scale.
Detailed gait and lameness research	Stride length, duty factor, joint angles, phase asymmetry, fine lameness grades	Side-view corridor, 2.0–2.5 m height, ≥1080p, 30–60 fps; fixed straight path	RGB for basic kinematics; stereo/RGB-D for 3D joint trajectories or vertical displacement	High-resolution 2D pose (HRNet, T-LEAP, DEKR), often with temporal encoders; optional multi-view 3D fusion; sequence classifier (RF, SVM, GB, CatBoost)	Prioritises distal joint fidelity and temporal consistency; higher annotation and calibration cost, but necessary when centimetre-level differences and phase curves are clinically relevant.
Body size, weight, and body condition	Body length/height, heart girth, surface area, BCS/BFT proxies, liveweight	Side or 3/4 view along alleyways; 1–3 m from animals; or overhead passways for cattle/pigs	Prefer RGB-D or stereo (Kinect, ZED, RealSense); RGB-only only if geometry is very controlled	2D keypoints (YOLOv8-Pose, Lite-HRNet, CowK-Net) projected to 3D point cloud; regression model for size/ weight	When centimetres matter, geometry beats texture: depth/ stereo reduces cross-farm domain shift and yields explainable Euclidean measurements. Best placed at routine traffic points (weighing, exit lanes).
Pen-level behaviour and welfare (activity, social interactions, tail posture, brush use)	Activity budgets, lying/ standing bouts, aggression events, tail posture, brush contact patterns	Overhead or high-angle pen views, 2–4 m height for cattle/pigs; 2–3 m for poultry	RGB sufficient for most behaviours; depth useful in very dense groups or complex housing	Bottom-up or top-down multi-animal pose (YOLOv8-Pose, HRNet, RTMPose, GRMPose) with tracking (SORT/BoT-SORT/ KeySORT); optional skeleton-GCN for interaction	Housing and density drive architecture choice: multi-animal bottom-up with tracking is preferable in crowded pens. Front-ends (cage-net removal, completion) help in poultry/ ducks with strong occlusion.
Identification, counting, and re-identification	Individual IDs, counts, time in zone, repeated visits	Top-down corridor or holding area, 2.5–3.5 m height; stable background	RGB; resolution more important than depth	Dorsal keypoints (10–12) + Siamese or template-based ID embeddings (SURABHI-style, bit-vector methods); lightweight detection (YOLOv8)	Works best in constrained zones (holding areas, ramps, parlours) where orientation is predictable. Periodic template refresh needed to handle growth and BCS changes.
Task-agnostic "future proof" sensing at high-value locations	Combination of mobility, morphometrics, and behaviour analytics	Combination of top-down + side-view at key chokepoints (corridors, parlours)	Stereo or RGB-D on at least one view; RGB on others	Modular design: robust 2D/ 2.5D pose front-end, pluggable downstream heads (lameness, weight, behaviour)	Higher upfront cost but maximises re-use: the same sensing rig can support multiple analytics over time as new models are trained, reducing disruption on farm.

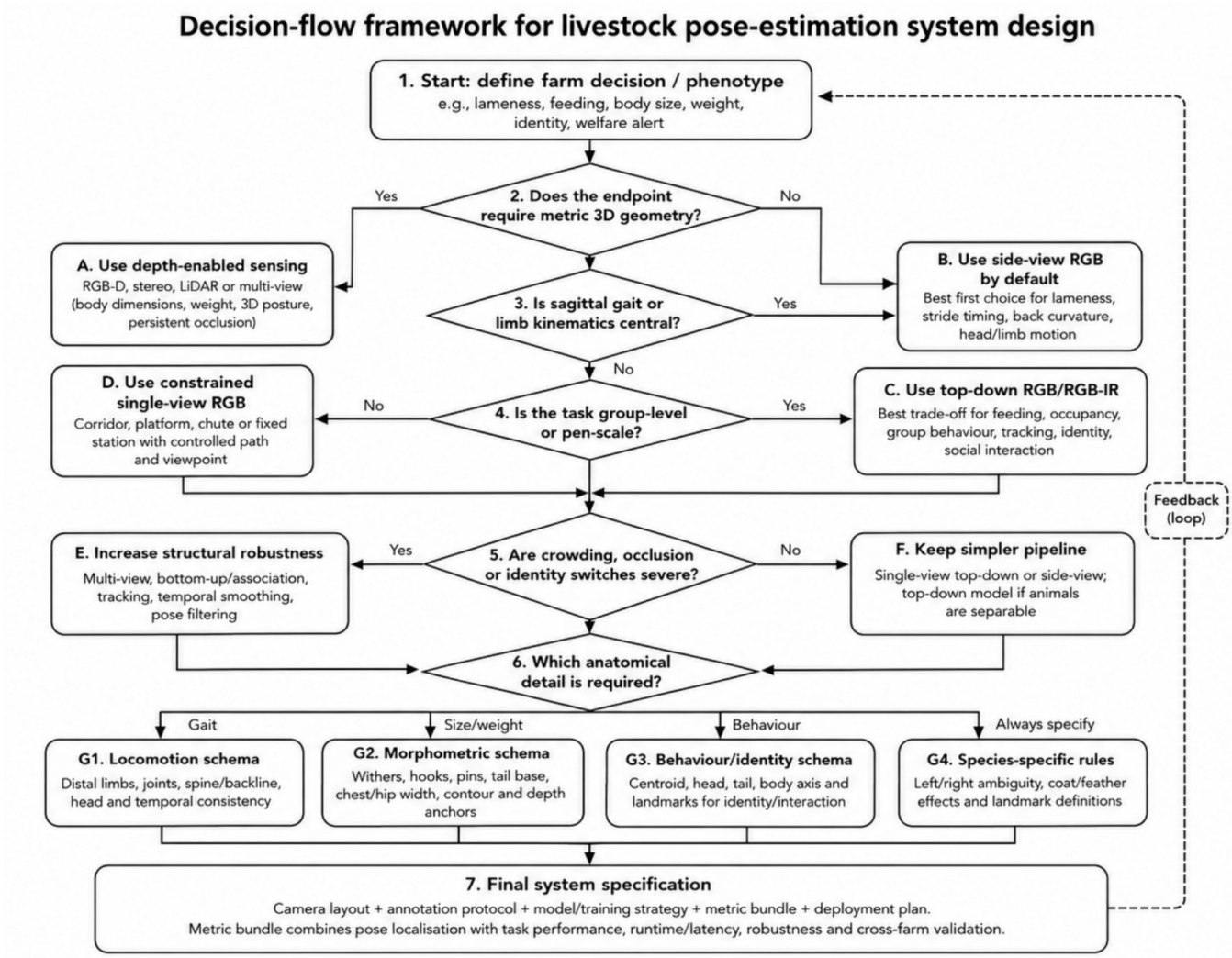

Fig. 15. Decision-flow framework for selecting livestock pose-estimation systems. The framework starts from the biological or farm-management task and progressively constrains the sensing modality, camera viewpoint, keypoint schema, model family, evaluation metrics, and deployment architecture. Its output is a complete system specification rather than a model choice alone.

body measurement, weight estimation, identity tracking, or welfare monitoring. The second input is the species and production context, including body morphology, coat or feather characteristics, animal density, housing system, available camera mounting locations, and whether the task occurs at a controlled checkpoint or in an open pen/pasture environment.

These inputs determine the sensing and viewpoint decision. Side-view RGB is the most appropriate default for sagittal gait, lameness, stride timing, and back-curvature analysis. Top-down RGB is generally preferable for group-level behaviour, feeding, occupancy, tracking, and identity-related applications. RGB-D, stereo, LiDAR, or multi-view systems are justified when the target output depends on metric geometry, body dimensions, persistent occlusion handling, or three-dimensional posture recovery. The next step is the keypoint-schema decision. Distal-joint-rich schemas are needed for detailed locomotor analysis, compact dorsal schemas are sufficient for many identity and activity-monitoring tasks, and measurement-oriented systems often require stable anatomical anchors combined with contour or depth-derived features.

The model choice should then follow from the animal density, viewpoint, and temporal requirements. Top-down methods are appropriate when individual animals can be reliably detected or separated. Bottom-up or association-based methods are preferable in crowded multi-animal scenes. Temporal, graph-based, or transformer-based models become important when the endpoint depends on movement continuity, social interaction, gait dynamics, or identity persistence. The evaluation layer should combine pose-localisation metrics with task-specific downstream metrics and deployment indicators. Finally, the deployment layer should specify hardware, latency, connectivity, calibration burden, maintenance requirements, user interface, and integration with farm-management actions. The recommended output of the framework is therefore not only a model, but a complete system specification: camera layout, keypoint definition, model family, training strategy, evaluation metric bundle, deployment hardware, and farm-decision pathway.

5. Conclusion

This systematic review synthesised 114 studies published between 2015 and 2025 on vision-based livestock pose estimation for precision livestock farming and showed that the field has evolved from isolated proof-of-concept studies toward increasingly deployment-aware systems for mobility analysis, lameness screening, behaviour recognition, morphometric measurement, weight estimation, welfare monitoring, tracking, and re-identification. Across species and applications, the central finding is that pose estimation is most effective when it is treated

not as an end in itself, but as an interpretable intermediate representation linking visual sensing to biologically meaningful livestock traits.

Three broad conclusions emerge from the updated evidence base. First, the selection of camera geometry is more decisive than the selection of any individual deep-learning architecture. Side-view RGB remains the most appropriate option for locomotion and lameness tasks requiring sagittal kinematics, whereas top-down RGB generally offers the best trade-off for group-level behaviour, feeding, social monitoring, and identity-related applications. Stereo, RGB-D, and multi-view systems become most valuable when the endpoint depends on absolute geometry, persistent occlusion handling, or three-dimensional posture reconstruction. Second, keypoint design must be task- and species-specific. Compact dorsal or trunk-centred schemas are often sufficient for monitoring, identification, and occupancy-related tasks, whereas detailed gait, lameness, and morphometric applications require denser landmark definitions, particularly around distal limbs or measurement anchors. Third, the main barriers to wider adoption are now less about raw model capability and more about dataset diversity, inconsistent annotation taxonomies, metric heterogeneity, and the limited reporting of deployment constraints, runtime assumptions, and farm-system integration.

The review also supports a practical decision pathway for selecting pose-estimation systems in livestock environments. When the target is behaviour, feeding, occupancy, or identity in group housing, a fixed top-down RGB setup should generally be considered first because it is scalable and cost-effective. When the target is gait, mobility, or lameness, side-view RGB with temporal modelling is usually the most suitable starting point. When the target is body size, body condition, or weight-related morphometrics, depth-enhanced or stereo systems are preferable because the required outputs are metric rather than relative. Multi-view systems should be reserved for scenarios in which single-view ambiguity, occlusion, or three-dimensional posture is intrinsic to the task and cannot be resolved reliably through simpler hardware.

At the same time, the review highlights important limitations in the current evidence base. Many studies remain single-site, single-breed, or tightly task-specific, which constrains confidence in cross-farm generalisation. Datasets are frequently private or only partially available, annotation protocols are rarely harmonised across groups, and evaluation metrics are often not directly comparable across studies. Long-term field validation, cost-effectiveness assessment, and explicit reporting of operational robustness remain uncommon relative to the large number of algorithm-development papers. These limitations mean that the field is advancing quickly, but still unevenly, with a gap between high benchmark performance and reproducible low-maintenance deployment at commercial scale.

Several priorities follow for future research. First, the community would benefit from species-specific core keypoint standards and task-aligned reporting bundles that combine pose accuracy with downstream performance and deployment indicators. Second, more multi-site, multi-breed, and openly described datasets are needed to support robust benchmarking, transfer learning, and external validation. Third, greater emphasis should be placed on annotation-efficient workflows, including active learning, semi-supervised learning, synthetic-data generation, and sim-to-real adaptation. Fourth, deployment studies should move beyond reporting nominal frame rates and instead include calibration burden, environmental robustness, uptime, edge-hardware performance, and integration with farm-management systems such as RFID-linked workflows or decision dashboards. Finally, emerging methods such as transformer-based architectures, self-supervised learning, federated learning, and richer multimodal fusion should be explored in ways that preserve anatomical interpretability and practical utility rather than adding complexity without clear farm-level value.

Finally, future studies should, where ethically and commercially feasible, release not only datasets but also annotation protocols, keypoint definitions, visibility rules, preprocessing scripts, evaluation splits, trained-model configurations, and code. When raw videos cannot be shared because of privacy, farm confidentiality, or commercial restrictions, authors should still provide detailed annotation guidelines, representative labelled frames, derived pose trajectories, metadata templates, and reproducible evaluation scripts. Such practices would substantially improve comparability, external validation, and reuse across species, farms, and camera configurations.

Overall, vision-based livestock pose estimation is now sufficiently mature to support a range of practical precision-livestock-farming applications, particularly in structured environments and well-defined workflows. The next stage of progress will depend less on increasingly complex backbones alone and more on standardisation, transparent reporting, robust cross-farm validation, and closer alignment between pose-derived analytics and real management decisions. With these advances, pose estimation has strong potential to become a core sensing layer for scalable, interpretable, and welfare-oriented livestock monitoring systems.

Funding

This paper has emanated from research funding provided by Science Foundation Ireland (SFI) which is Co-Funded through the European Regional Development Fund under Grant 12/RC/2289-P2-INSIGHT, and the SFI Centre VistaMilk (SFI 21/RC/10303).

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used Copilot in order to check the language and improve clarity of the text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

CRediT authorship contribution statement

Matteo Menolotto: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Salvatore Tedesco:** Writing – review & editing, Validation, Methodology. **Luke O'Grady:** Writing – review & editing, Validation. **Emer Kennedy:** Writing – review & editing. **Brendan O'Flynn:** Writing – review & editing, Supervision, Resources, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix 1. Extensive data extraction from the selected articles

No.	Work	Livestock targeted	Camera technology	Camera location	ML/DL models used	Pretrained networks used	Dataset used	Performance obtained
1	Liu et al. (2025a)	Pigs	Overhead RGB cameras	Overhead	MT-SRNet, HRNet-based multitask network, SimCC	HRNet variants	PIG10, PIG-ALL	mPCK 95.33%; mIoU 95.52%; posture accuracy 89.72%; up to ~ 12,000 FPS; ~30 × fewer parameters
2	Chen et al. (2025c)	Sows	GW800 RGB camera	Side view	FYOLOv8s-Pose, NLPoseC3D, SimTSM, fusion classifier	No standard pretrained model stated; custom YOLOv8s-derived backbone	Custom sow dataset, 459 videos, 2,745 labeled images	Pose mAP 95.9%; fusion accuracy 94.2%; sensitivity 97.4%; specificity 96.8%
3	Nasiri et al. (2022)	Broiler chickens	Intel RealSense D455 (RGB mode)	Side, angled 30°	DeepLabCut with ResNet-50, LSTM	ResNet-50 pretrained on ImageNet	Broiler Pose Dataset, 9,214 images, 400 videos	Accuracy 97.5%; precision 95.0%; sensitivity 95.45%; specificity 98.39%; AUC 96.92%
4	Xu et al. (2025b)	Pigs	Microsoft Kinect DK depth cameras	Arch-shaped multi-view, top + side	PointNet++, KPConv, RS-CNN, SparseConv3d	Not specified	937 pig point clouds from 148 pigs	Keypoint RMSE 4.115 cm; body-size MAPE 3.90%
5	Luo et al. (2023)	Cattle and pigs	Microsoft Kinect v2 RGB-D	Left, right, top	SMAL, DeepLabCut, non-rigid ICP, PCA, L-BFGS-B	SMAL trained on animal toys; ResNet pretrained on ImageNet	103 cattle and 13 pigs, RGB-D + point clouds	Cattle accuracy 91.95%; pig accuracy 87.63%; pose normalisation improved measurement accuracy
6	Farahnakian et al. (2024)	Sows and piglets	Dahua IP cameras	Overhead	DeepLabCut models with ResNet-50, ResNet-101, MobileNet, EfficientNet, DLCRNet	ImageNet via DeepLabCut	856 annotated frames from 310 h, 16 farrowings	Best median test error 0.61 px with MobileNet; DLCRNet RMSE 0.39–0.95 px
7	Kroese et al. (2025)	Dairy cows	Ubiquiti G3 Bullet RGB cameras	Overhead + side	HRNet, Random Forest, RNN, phase detection pipeline	HRNet	809 rising + 791 lying sequences from 183 cows	MAD < 0.5 s for most phases; worst ~ 0.9 s; 59.9% of rising and 29.1% of lying bouts exceeded comfort thresholds
8	Gan et al. (2021)	Sows and piglets	Hikvision DS-2CD1321D-I	Overhead	ResNet-18, ResNet-50, FlowNet2.0, CNN behaviour network	FlowNet2.0 pretrained on synthetic data	11.7 h train + 24.7 h test video	Accuracy 97.63%; sensitivity 98.55%; specificity 97.36%; 6.36 fps
9	Wang et al. (2023b)	Pigs	Hikvision smart265	Overhead	HRNetV2, GAN, DARK, PAF	HRNetV2 pretrained	1,780 images from 40 videos of 14 pigs	mAP 74.09%; AP75 74.30%; PCKh improved by 11.37% over baseline
10	Lodkaew et al. (2023)	Cattle	FireFly MV CMOS USB camera	Overhead	YOLOv4, ResNet-50, DenseNet-169, EfficientNet-B0, XGBoost, CatBoost, LightGBM	ImageNet for CNNs; COCO for YOLOv4	2,000 labeled images, 3 cows	Detection mAP 99%; body-part MAE 4.3 px; estrus F1 91% cow-specific, 80% end-to-end
11	Yu et al. (2024)	Poultry	Azure Kinect DK	Robot-mounted, adjustable	CCMNN / CycleGAN, YOLOv8 variants	YOLOv8 pretrained on COCO; CNNs on ImageNet	21,600 RGB images, 13,884 retained	SSIM 91.14%; PSNR 25.34 dB; precision improved by up to 11.1%
12	Taghavi et al. (2024)	Dairy cows	Hikvision DS-2CD2T45FWD-I5 RGB	Side view	T-LEAP temporal encoder-decoder CNN	T-LEAP pretrained on outdoor cow dataset	758 samples from 44 clips, 34 cows	Overall PCKh@0.2 89%; legs 96%; head 98%; back 59%
13	Nasir et al. (2025)	Cattle	Hikvision 4 K cameras	Multi-angle overhead	YOLOv5l, YOLOv8l, MMIU, BEV fusion	Ultralytics pretrained YOLO models	67,840 boxes, 4,188 keypoints, 6,398 frames	Action detection mAP 99.2%; pose improved with YOLOv8; fusion errors reduced with SPM
14	Siachos et al. (2024)	Holstein dairy cows	2D surveillance camera	Overhead	EfficientNetV2 ordinal regression, RMSprop	EfficientNetV2 pretrained	34,150 training BCS; 9,657 testing BCS	Weighted kappa 0.69 for single BCS; PA within ± 0.25 = 84.6%; Spearman correlation with backfat 0.75
15	Martin-Cirera et al. (2024)	Horses	GoPro Hero 4	Overhead	HRNet, LSTM, BiLSTM, Transformer	HRNet pretrained on COCO	6,000 keypoint images; 10 videos; 193,140 frames	HRNet PCK@0.2 94.17%; best F-score up to 0.94
16	Fang et al. (2021)	Broiler chickens	Logitech C922 webcam	Overhead	DeepLabCut with ResNet-50, Naive Bayes	ResNet-50 pretrained on ImageNet	B6 dataset, 28 videos, 556 frames	Accuracy 0.781; F1 eating 0.944; resting 0.954; walking 0.532

(continued on next page)

(*continued*)

No.	Work	Livestock targeted	Camera technology	Camera location	ML/DL models used	Pretrained networks used	Dataset used	Performance obtained
17	Fan et al. (2023)	Dairy and beef cattle	RGB cameras, unspecified	Mixed farm scenes	CMBN, DO-Conv, CoordConv, Triplet Attention	HRNet pretrained	NWAFU-Cattle; Kaggle dairy cattle dataset	AP 93.2; AR 96.4; 14.01 M params; 9.83G FLOPs
18	Bai et al. (2025)	Cattle	Huawei Nova11 phone camera	Side + rear	MobilePoseNet, PoseNet, LaWE regression network	MobileNetV3 Small	108 cattle; 2,550 samples	RMSE 13.1–15.4 kg; MAPE 3.3–3.7%; R^2 0.886–0.897; mAP50 86.2%
19	Högberg et al. (2025)	Dairy cows	Ubiquiti G3 Bullet RGB cameras	Top-down, side, high-angle	YOLOX, HRNet, Random Forest, LSTM	HRNet retrained on proprietary cow images	25.5 h video; 1,777,949 frames; 222 bouts	Brush-use accuracy 86.3%; segment detection 87% with spline method; cow ID 98%
20	Li et al. (2024c)	Simmental beef cattle	Wide-angle RGB camera	Side view	YOLOv5s, Lite-HRNet, Global-Local Path networks	Custom-trained on beef cattle images	2,956 images of 30 cattle	AP 97.1%; OKS 0.918; avg measurement error 6.75–8.97%
21	Fodor et al. (2025)	Broiler chickens	Reolink RLC-510A	Overhead	YOLOv8 pose	YOLOv8 pretrained, fine-tuned	Custom corridor dataset, 118 broilers	mAP50-95 0.81 for boxes; 0.99 for head keypoint; F1 0.99 at IoU 0.5
22	Yang et al. (2025)	Dairy cows	ZED stereo camera	Side-mounted	CowK-Net with ResNet-50 backbone, hybrid encoder, Transformer decoder	ImageNet pretrained ResNet-50	3,512 RGB-D images; CowDatabase, CowDatabase2	AP_kp 92.8%; mean relative errors: BL 2.8%, BH 6.7%, HH 4.1%, CD 4.4%
23	Liu et al. (2023d)	Multiple incl. cattle, poultry, sheep, horses	RGB cameras, unspecified	Not specified	MAPoseNet base and large	No explicit pretraining stated	AP-10 K; Animal Pose	AP 74.3 val / 72.6 test; AR 77.4 val / 75.8 test; outperformed HRFormer
24	Liu et al. (2020)	Dairy cows	Surveillance, GoPro, IP cameras	Side view	DeepLabCut, custom CNNs	ImageNet pretraining, then fine-tuning	3 datasets, 240 train frames, 1,255 test frames	Body F1 up to 0.879; leg-hoof F1 up to 0.434; VCP up to 0.960
25	Higaki et al. (2025)	Dairy cows	GoPro Hero9	Overhead	AiPEC + Random Forest	Proprietary AiPEC; pretrained not stated	256 cows, 256 clips	Weighted kappa 0.72; AUC-ROC 0.89; class accuracies 92.1%, 76.0%, 82.3%
26	Wu et al. (2024)	Dairy cows	Canon EOS 5D Mark III	Side view	ResNet101-ASPP, DeepLabCut; compared with ResNet50, MobileNet, EfficientNet	ResNet pretrained on ImageNet	238 videos of 108 cows	RMSE 6.98 px; confidence up to 96.79%
27	Chen et al. (2025c)	Piglets	Hikvision DS-2CD3325D-I	Top-down	YOLOv11s-Pose, BoT-SORT	YOLOv11s pretrained	1,640 detection images; tracking data from 6 pens	AP@0.5 98.5% objects, 98.0% keypoints; HOTA 84.3%; MOTA 99.1%; IDF1 91.5%
28	Lu et al. (2025)	Cattle	Microsoft Kinect v2 RGB-D	Left, right, top	DeepLabCut, SMAL, encoder-decoder GCN	DeepLabCut pretrained; custom GCN	100 cattle RGB-D + point clouds	Average MAPE 3.58% across 9 traits; Chamfer distance 4.01 cm; runtime ~ 19 s
29	Zhao et al. (2025a)	Ducks	IMX577 camera modules	Robot-mounted, cage layers	Mask R-CNN, CycleGAN, YOLOv7-CRD, HRNet-48	COCO pretrained weights	Multiple datasets from 10-, 20-, 30-day ducks	OKS improved 13.22–18.07%; mAP improved 7.48–10.35% after image completion
30	Guan and Chai (2025)	Pigs	Logitech C920 Pro webcam	45° downward view	YOLOX, HRNet, GCN, RepVGG	AnimalPose, AP-10 K, ATRW	7,858 RGB images; 2,104 skeleton samples	Pose AP 96.8%; action accuracy 94.3% skeleton, 99.4% RGB
31	Perneel et al. (2025a)	Dairy cattle	Hikvision DS2CD2643G2-IZS	Roof-mounted, top-down	Custom separable-conv CNN, KeySORT	None, trained from scratch	500 annotated frames; 100 test clips	Keypoint recovery 78.4% direct, 79.9% with KeySORT; precision up to 89%; recall up to 66%
32	Yue et al. (2025)	Dairy cattle	Kinect V2	Side-mounted	YOLOv8-Pose-LSCDH; compared with HRNet, DEKR, YOLOv5s/7-Pose	YOLOv8 pretrained weights	Custom dataset + Animal-Pose	mAP50-95 up to 96.1%; body measurement error < 4%
33	He et al. (2025)	Pre-weaning piglets	Infrared IP dome camera	Overhead	YOLOv8s, YOLOv8s-LSKNet, YOLOv8s-ELSKNet, RTMPose + RTMDet, YOLOv5s	Ultralytics YOLOv8 pretrained	2,000 images from 3 h video	Pose mAP@0.5:0.95 95.3%; box mAP 86.1%; LOF filter precision 96.8%

(*continued on next page*)

(*continued*)

No.	Work	Livestock targeted	Camera technology	Camera location	ML/DL models used	Pretrained networks used	Dataset used	Performance obtained
34	Deng et al. (2025)	Cattle	ZED stereo camera	Side view	MobileViT-Pose, LiteMobileViT-Pose, YOLOv8n, CREStereo	ETH3D for CREStereo; ImageNet for MobileViT	NWAFU-Cattle; CowDatabase; CowDatabase2	Pose AP 92.4%; body-size error 2.85%; speed 18.8 fps
35	Qu et al. (2025)	Goats	Hikvision network camera	Overhead	YOLOv8s-pose, ByteTrack, Bayesian optimisation	Ultralytics YOLOv8s-pose pretrained	1,466 labeled images + 464 clips	Accuracy 95.04%; recall 96.55%; FPR 6.47%; ~0.09 s/frame
36	Tan et al. (2025)	Pigs	Orbbec Astra Pro Plus RGB-D + Hikvision binocular camera	Overhead	RTMDet, RTMPose, CatBoost, XGBoost, Random Forest	RTMPose via MMPose; RTMDet pretrained	17,406 segmentation images; 6,048 RGB-D images	R^2 0.82; MAPE 3.63%; AP bbox 94.69%; AP keypoints 90.59%
37	Liu et al. (2023b)	Dairy cattle	Ubiquiti UniFi UVC-G3-FLEX cameras	Side + front	DeepLabCut, Mask R-CNN, Random Forest, SVR, Linear Regression	Mask R-CNN pretrained on COCO; DeepLabCut pretrained on Purdue data	505 samples from 196 cows	RMSE 34 kg; MAPE 4.1%; 57% within 25 kg
38	Russello et al. (2022)	Dairy cows	ZED RGB-D stereo camera, RGB only used	Side view	LEAP, T-LEAP	None, trained from scratch	CoWalk-10 and CoWalk-30, 1,059 samples	PCKh@0.2 99% clean, 88.4% occluded, 87.6% unseen cows
39	Li et al. (2022)	Dairy cows	Hikvision dome cameras	Side view	ResNet101, ST-GCN, VGG-19, multi-branch CNN fusion	No pretrained weights stated	680 videos	Accuracy 97.20%; TPR 97.05%; FPR 1.21%
40	Higaki et al. (2024)	Dairy cows	GoPro Hero 9	Side view	AiPEC + Random Forest	None stated; AiPEC proprietary	204 cows, 204 walking videos	Weighted kappa 0.69; AUC-ROC 0.86; accuracy 83%; F1 74–76%
41	Chen et al. (2024)	Dairy goats	Infrared network cameras	Overhead	GRMPose with CSPNext + GCNCC; baselines HRNet, ResNext, MobileNetv2, ShuffleNetv2, SimpleBaseline, Pvtv2	CSPNext pretrained via MMPose	DairyGoat dataset, 2,108 train, 258 test images	AP 87.48%; latency 13.82 ms; 5.57 GFLOPs; 27.56 M params
42	Kroese et al. (2024a)	Dairy cows	Ubiquiti G3 Bullet	Side-mounted multi-view	CNN-based 2D detector and pose estimator, 3D pose pipeline	Not explicitly stated	979 sequences from 183 cows	Sensitivity 88.5%; precision 99.5%; ICC 0.81–0.85
43	Menezes et al. (2025)	Dairy cows	GoPro Hero 9 + Intel RealSense D435	Top-down	Stacked hourglass pose net + Siamese neural network with triplet loss	Not specified	4,319 images for pose; 188 videos for identification	Closed-set accuracy up to 96.3%; open-set 75.4%; perturbation tolerance 2% with F1 78.4%
44	Fodor et al. (2023)	Broiler chickens	Intel RealSense D415	Rear view	DeepLabCut-based CNN pose estimator	Pretrained DeepLabCut model retrained	Videos of 84 broilers at 3 ages	Train error 2.11 px; test error 4.02 px
45	Zhao et al. (2023a)	Dairy cows	Nikon D5200	Side view	DeepLabCut, Logistic Regression, SVM, LDA, Naive Bayes, Decision Tree, KNN, FFNN	ResNet50 pretrained on ImageNet	670 videos, 410 used	Accuracy 89.2%; recall 90.7%; precision 88.2%; specificity 87.6%
46	Du et al. (2022)	Cattle and pigs	Microsoft Kinect v2 RGB-D	Left, right, top	DeepLabCut with ResNet backbone	ResNet pretrained on ImageNet	103 cattle and 13 pigs scanned with 3 cameras	Most measurement MAPE < 10%; keypoint RMSE < 8 px
47	Rozendo et al. (2025)	Cattle	Standard RGB camera	Side view	MMPose, CAFE-Net, Depth Pro, LASSO, RF, ANN	ResNet50 for MMPose; pretrained Depth Pro	Kaggle B3 and B4 datasets	Best MAPE 10.02% after filtering; MAE ~ 17.55 kg
48	Jia et al. (2025)	Dairy cows	Sony FDR-AX45	Side view	DeepLabCut with dlcrnet_ms5	ResNet-50, ResNet-101, MobileNetV2, EfficientNetV2 backbones tested	127 videos from 45 cows	Accuracy 90.21%; keypoint error 4.679 px
49	Li et al. (2024a)	Horses	Synthetic PyTorch3D rendering / monocular RGB	Monocular views	DINO-ViTs8, HMR-style modules, custom decoders, hSMAL	DINO-ViTs8 pretrained	DessiePIPE; PFERD; MagicPony; Pascal; AnimalPose	PCK@0.1 up to 75.43%; IoU up to 59.99%; Chamfer distance 11.85 mm
50	Guo et al. (2025)	Pigs	RGB overhead camera, unspecified	Overhead	YOLOv8x-pose-p6, ResNet-18, MLP	COCO-pretrained YOLOv8x-pose-p6	Norwegian and German pig datasets	PDJ 48.96%; PCK 84.85%; pig OKS 89.43%; pose accuracy 99.33%
51	Ramesh and Reibman (2024)	Holstein cows	Top-view camera, unspecified	Overhead	KeypointRCNN + SURABHI self-training	ImageNet pretrained weights	Purdue Dairy cow videos; 717 annotated images; 148 cows	Top-6 accuracy 91.89%; 22% more correct IDs than baseline

(*continued on next page*)

(*continued*)

No.	Work	Livestock targeted	Camera technology	Camera location	ML/DL models used	Pretrained networks used	Dataset used	Performance obtained
52	Yang et al. (2024)	Angus beef cattle	Intel RealSense D435i	Side view	YOLOv8-SimBiFPN	YOLOv8	1,905 images, 2,401 instances; 58 cattle measured	mAP@0.5 99.1%; MAPE 4.14–6.47%; MAE 0.039–0.055 *m*
53	Peng et al. (2024)	Simmental beef cattle	ZED2I depth camera	Side view	Improved YOLOv8-pose with SimSPPF and CARAFE	YOLOv8	2,985 images; 23 cattle measured	mAP 94.4%; MAE 1.52–9.77 cm; MRE 1.28–6.47%
54	Zhao et al. (2024)	Dairy and beef cattle	iPhone 12	Farm-level, natural habitat	YOLOv8n, SimAM, EfficientRepBiPAN	YOLOv8n pretrained	6,846 images, 8,600 cattle instances	AP@0.5 92.3%; AP@0.5:0.95 56.4%; 3.11 M params; 8.1G FLOPs
55	Russello et al. (2024)	Dairy cows	ZED RGB-D stereo camera	Side view	T-LEAP, Faster R-CNN, SVM-R, RF, GB, MLP	Faster R-CNN pretrained on COCO-2017	272 videos, 98 cows; 28 annotated videos for pose	PCKh@0.2 99.6%; best lameness accuracy 80.07%; F1 78.70%
56	Su et al. (2024)	Mongolian horses	ORDRO V17	Side view	SimAM-YOLOv8n, RTMPose	RTMPose pretrained on MMPose	4,000 augmented images; 20 horses test	AP-obj 90.1%; AP-kp 91.4%; MRE 2.74–5.28%
57	Fang et al. (2024)	Chickens	Logitech C922	Overhead	YOLOX + EfficientNet	EfficientNet pretrained on ImageNet	3,831 images from 29 videos	mAP 0.652; mAR 0.742; PCK 0.789; test RMSE 17.30 px
58	Li et al. (2024b)	Cattle	Snapdragon S3 sports camera	Handheld, multi-angle	Modified YOLOv8-pose with SimSPPF, ECA, RepGFPN	Not specified	2,034 annotated images, 12 cows	mAP 96%; chewing count error 5.6%; precision 96.8%; recall 87.2%
59	Okour et al. (2024)	Cattle	Multi-depth camera system	Overhead + side	PointNet++, MLP	Not specified	Synthetic dataset + 175 real cattle for validation	Joint error 0.026 m mean; hip-height RMSE 2.97; R^2 0.64
60	Zhang et al. (2024)	Sheep	Xiaomi Smart Camera 3	Side-mounted	RTMPose, RTMDet	SimCC-based classification; pretraining not specified	20 videos; 4,375 boxes; 26,250 keypoints	Chewing rate 97.99%; rumination time 97.49%; number of ruminations 94.56%
61	Liu et al. (2024)	Camels	Canon 60D	Side + varied angles	YOLOv11pose with ECA, SimSPPF, DECA-block	Pretrained on ImageNet	1,084 annotated images	mAP@0.5 94.1%; precision 94.5%; recall 94.3%; 147.6 FPS
62	Kroese et al. (2024b)	Dairy cows	Ubiquiti G3 Bullet cameras	Overhead	Custom 3D pose pipeline	Not specified	493 rising bouts from 183 cows; 165 annotated bouts	ICC 0.95 machine vs human; MAD 0.33 s; mean lunge angle 166.1°
63	Li et al. (2023)	Beef cattle	Monocular RGB camera	Overhead + angled views	Improved RTMPose-FSD with FasterNest, SimAM, Depth Block	RTMPose / COCO pretrained weights	3,036 custom cattle images + 3,000 + AP-10 K	82.9% AP on cattle; 72.0% AP on AP-10 K; 39 FPS
64	Zhao et al. (2023b)	Ducks	Intel RealSense D435i RGB	Side, angled	HRNet-32 with CBAM	HRNet-32 pretrained on animal pose data	3,800 duck images + 300 duckling images	AP 0.943; model size < 25 MB; real-time inference demonstrated
65	Wei et al. (2023)	Dairy cows	5 MP infrared camera	Overhead	Multi-stage CNN with VGG-19 + MS-TCN	VGG-19 pretrained on COCO 2017	3,900 labeled images + 1,800 temporal sequences	APK for pelvis 90.13%; MS-TCN precision 94.71%; recall 86.99%; F1 90.69%
66	Liu et al. (2023c)	Pigs	Hikvision binocular RGB camera	Overhead	Mask R-CNN, Keypoint R-CNN, improved ResNet regressor	COCO pretraining for Mask R-CNN	6,384 pig images from 117 pigs	Segmentation AP 91.04%; keypoint AP 89.59%; weight RMSE 3.52 kg
67	Weng et al. (2023)	Cows	Hikvision IR + wide-angle RGB cameras	Overhead + side	Improved U-Net, Hungarian matching, Flow Shift + Kalman	U-Net, no explicit pretraining stated	~3,000 labeled frames from ~ 300 videos	OKS 0.92 mean; mAP 0.88; mAR 0.89; Error@95% 7.64 px
68	Lawin et al. (2023)	Horses	iPhone 12 Pro Max	Front-facing	Custom deep neural networks for keypoint tracking	Not described	25 horses with synchronized motion capture	Mean difference < 2.2 mm; RMSD 5.0 mm head, 3.5 mm pelvis
69	Ong et al. (2023)	Cattle	Fixed RGB CCTV, Hikvision	Overhead	HRNet, YOLOv8-Pose, BoxMOT, YOLOv5-Seg, RT-DETR	YOLOv8 pretrained on COCO; HRNet trained on AP-10 K	30,703 annotated frames; 753 cattle instances	YOLOv8-Pose OKS 0.638; HRNet OKS 0.459; detection mAP0.5 up to 0.956
70	Ramesh et al. (2023)	Holstein cattle	Ubiquiti UniFi UVC-G3 RGB	Overhead	Keypoint R-CNN with ResNet50-FPN / ResNet101-FPN	COCO-pretrained Keypoint R-CNN	Purdue Dairy Cow Keypoints Dataset; Cow Images Dataset	Top-1 accuracy 61.5%; Top-4 83.1%; R50-FPN better than R101-FPN
71	Yang et al. (2023)	Cattle	Multiple RGB cameras	Mixed angles	YOLOX, HRNet, ST-GCN, SCIRNet three-stream fusion	HRNet pretrained on AP-10 K	100 short videos with interaction labels	SCIRNet accuracy 75.2%; higher than

(*continued on next page*)

(*continued*)

No.	Work	Livestock targeted	Camera technology	Camera location	ML/DL models used	Pretrained networks used	Dataset used	Performance obtained
								ST-GCN 58.9% and AIGCN 72.7%
72	Liu et al. (2023a)	Piglets	Hikvision smart ball camera	Overhead	Modified DeepLabCut, ZS thinning, SVM, PAF	ResNet-50 pretrained on ImageNet	5,000 videos; ~14,000 train images	Node tracking accuracy 85.08%; posture classification 89.60%
73	Niknejad et al. (2023)	Horses	ZED2 stereo RGB camera	Side view	DeepLabCut with ResNet101, Faster R-CNN, SGBM	ImageNet for DLC; COCO for Faster R-CNN	40 stereo horse videos	Head-length RMSE 0.028 m; stance-duration RMSE 0.094 s; ICC 0.79 stance, 0.98 stride
74	Gorssen et al. (2022)	Pigs	Dahua IPC-HDW4831EMP-ASE	Overhead	DeepLabCut with ResNet-50; genetic REML analysis	ImageNet pretrained ResNet-50	7,428 annotated records from 1,556 pigs	Tracking RMSE ~ 3.3 cm; tail-neck r = 0.94; hip-width r = 0.80; heritability 22–74% depending on trait
75	Feuser et al. (2022)	Horses	iPhone 11	Side view	DeepLabCut with ResNet-50	ImageNet pretrained ResNet-50	215 videos; 22 lame and 8 sound horses	Forelimb SE/SP 100%; hindlimb with stifle ~ 89% SE/SP
76	Wang et al. (2022)	Pigs	Hikvision DS-2CD3346FWDA3-1	Overhead	HRST, CenterNet, Swin Transformer	ImageNet pretrained HRNet backbone	18,191 posture images; 1,623 keypoint images	AP 77.4%; 40 FPS; parameters reduced 72.8%; FLOPs reduced 41.7%
77	Gong et al. (2022)	Dairy cows	Hikvision 5 MP infrared camera	Fence-mounted, 4 m high	YOLOv4, two-branch PAF CNN, FCNN	COCO2017 for pose; PASCAL VOC 2012 for detection	175 videos; 1,800 train images; 400 test	Keypoint AP up to 91.6%; pose classification accuracy 94.62%; detection AP up to 97.61%
78	Fang et al. (2022)	Broiler chickens	1080p HD camera	Overhead	Faster R-CNN with ResNet-101	COCO-pretrained ResNet-101	Custom poultry dataset, size not clearly stated	AP per part 89–95%
79	Yang et al. (2022)	Pigs	Industrial camera, 4 mm lens	45° and 90° overhead angles	Custom CNN with Triplet Attention and deconv head	No pretrained backbone stated	1,000 images, 5,236 pigs	AP 93.4%; AP50 96.5%; AP75 95.5%; AR 95.0%
80	Khin et al. (2022)	Cattle	Surveillance cameras, unspecified	Side + diagonal	DeepLabCut with ResNet-50, SVM	ResNet-50 pretrained on ImageNet	14 videos, ~4,200 frames	Average accuracy 88.75%; pose-specific precision 84–95%; recall 78–96%
81	Kresovic et al. (2022)	Pigs	RGB camera 2688 × 1520, 15 fps	Overhead / slightly angled	OpenPose-style multi-stage CNN, EfficientNet-B6, CBAM	EfficientNet-B3/B3a/B6 pretrained	870 training + 217 validation images	Best performance with EfficientNet-B6; inference time ~ 0.8 s/frame
82	Farahnakian et al. (2021)	Pigs	Kinect v2 + Lorex cameras	Overhead	DeepLabCut with ResNet-50	ImageNet pretrained ResNet-50	2,000 annotated images, 21,966 pigs	RMSE 10.14 seen, 12.02 unseen; training RMSE 6.70
83	Yik et al. (2020)	Pigs	Intel RealSense D435, IR + depth	Overhead	Stacked Hourglass Network	None, trained from scratch	DIAT dataset, 6,000 labeled frames	Avg keypoint error ~ 9.7 px; miss rate < 5%
84	Li et al. (2019)	Pigs	RGB surveillance cameras	Overhead	Stacked Hourglass Network, Kalman filter tracking	MPII Human Pose pretraining for SHN	1,000 training images; 1,800 test frames	PCK > 90% for head and tail; tracking accuracy ~ 5% better than bbox methods
85	Hewitt and Mahmoud (2019)	Sheep	RGB still images and frames	Frontal / varied close-up	ERT, PI-ERT, Hopenet	Hopenet pretrained on 300 W-LP	SFLW, 850 sheep-face images	MNE improved 0.054 → 0.045; success rate 0.88 → 0.94
86	Volkmann et al. (2022)	Turkeys	AXIS M1125-E RGB IP camera	Overhead	HRNet-W48, ResNet-50/101/152, U-Net	HRNet-W48 and ResNet pretrained on ImageNet	244 images, 7,660 turkeys	Best model HRNet-W48 AP50 = 0.74
87	Riekert et al. (2020)	Pigs	Hikvision IP dome camera	Overhead / high-angle	Faster R-CNN with NAS backbone	NAS pretrained on MS COCO	7,277 annotations from 305 images across 31 videos	mAP up to 80.2% seen pens; 44.8–58.8% unseen pens; AP for position > 87% seen
88	Li et al. (2019)	Cattle	iPhone 8 Plus	Handheld, varied angles	CPMs, Stacked Hourglass, Heatmap Regression	VGG-FCN for heatmap regression	2,134 images of 63 cattle	Mean PCKh: stacked hourglass 90.39%; CPMs 88.29%; heatmap regression 83.52%

(*continued on next page*)

(*continued*)

No.	Work	Livestock targeted	Camera technology	Camera location	ML/DL models used	Pretrained networks used	Dataset used	Performance obtained
89	Witte et al. (2024)	Pigs	AXIS M3206-LVE and M3116-LVE	Overhead	YOLOv8s, EfficientNetV2-B0	EfficientNetV2 pretrained on ImageNet	9,218 images for pig detection; 90,048 posture; 18,289 tail posture	Pig detection mAP@0.5:0.95 0.957; tail-posture mAP@0.5:0.95 0.601; posture F1 0.97
90	Barney et al. (2023)	Dairy cows	GoPro Hero6	Side view	Mask-RCNN, SORT, CatBoost	ResNet-101 backbone for Mask-RCNN	25 videos, 250 cows + 500 Google images	Lameness accuracy 98%; severity classification 94%; kappa 0.8782
91	Perneel et al. (2025b)	Holstein-Friesian cattle	Hikvision DS-2CD2643G2-IZS	Overhead	Mask-RCNN, custom CNN, k-NN, CatBoost	Benchmarked against ImageNet-pretrained AlexNet, ResNet-50, InceptionV3	11,438 annotated segments from 48 cattle	Rank-1 accuracy 0.894 with behaviour/ orientation correction; 0.822 without
92	Klecel et al. (2025)	Cattle and horses	Standard video files	Side view	DeepLabCut with ResNet-50, refineDLC post-processing	Custom-trained ResNet-50	17 cattle videos; 524 horse videos; 2,268 labeled frames	Reduced noise and improved physiological plausibility; validated filtering thresholds
93	Shirey et al. (2023)	Horses, adaptable to other livestock	Not specified	Not specified	DeepLabCut-based analysis GUI	Transfer learning via DeepLabCut	Horse locomotion videos	Functional GUI for synchronized video + kinematic analysis; exports gait parameters
94	Narli et al. (2025)	Dairy cattle	Intel RealSense D435 RGB-D	Side view	DeepLabCut with ResNet-152, fully connected DNN	ResNet-152 backbone, pretraining not stated	510 videos from 260 cows; 7,650 annotated frames	Non-lame vs lame accuracy 0.97; LCS1 vs 2 accuracy 0.88; LCS3 vs 4 accuracy 0.93
95	Liu et al. (2025b)	Yak heifers	Hikvision DS-2CD3T47EWDV3-L	Overhead multi-camera	MFPVT-YakHeifer with PVT v2, MFFM, DS-CBAM, LowerDeconv	Object detector pretrained; other pretraining not explicit	YakPoseData, 1,205 images, 3,223 yaks	mAP0.5 90.41%; mAP 64.11%; mAR0.5 91.95%; 7.00 ms on Ubuntu; 248.10 ms on Jetson Nano
96	Qu et al. (2025)	Goats	Hikvision network camera	Overhead	YOLOv8s-pose, ByteTrack, Bayesian optimisation	Ultralytics implementation; explicit pretraining not stated	1,466 labeled images + 464 clips	Head-to-head recall 97.32%; head-to-body recall 95.18%; overall accuracy 95.04%; ~24 FPS
97	Högberg et al. (2025)	Dairy cows	Ubiquiti G3 Bullet RGB cameras	Top-down, side, high-angle	YOLOX, HRNet, Random Forest, LSTM, spline-based segment inference	Retrained on proprietary cow images	25.5 h video, 1,777,949 frames	Brush-use accuracy 86.3% RF, 83.9% LSTM; segment macro accuracy up to 88%; brush-user ID 98%
98	Liu et al. (2025a)	Pigs	Overhead RGB cameras	Overhead / top-down	RBB detector, MT-SRnet on HRNet, SimCC, mask head, posture classifier	HRNet backbone, pretraining not stated	PIG-ALL; PIG10-HBB/RBB/MASK; PIG10-TEST	mPCK 95.33%; mIoU ~ 95.53%; posture accuracy 89.72%; ~17,000 pigs/s; ~248 FPS pipeline
99	Wu et al. (2025)	Holstein dairy cows	Canon EOS 5D Mark III	Ground-level side view	Improved YOLOv8n, improved RTMPose, Random Forest; compared with SVM, LR, KNN, XGBoost	COCO + AIC pretraining for pose	Dairy-cow-dataset: 784 clips, 2,827 annotated frames; lameness set 418 cows	Detection AP50-95 up to 94.8%; pose mAP 95.8%, mAR 96.3, 147.5 FPS; lameness precision / accuracy 93.8%
100	Wei et al. (2025)	Sheep	Kinect V2 depth cameras	Top + left + right	PointNet++, FPFH + RANSAC + ICP, geometric post-processing	None stated	35 Dorper sheep; subset of 24 for evaluation	Segmentation best accuracy 0.894 / 0.912; mIoU up to 0.753; measurement errors 1.14–3.71%
101	Hu et al. (2025)	Group-housed piglets	Hikvision DS-2CD3345D-I	Overhead	Pig-HRNet, K-means, hyperplane classifier; compared with HRNet, DLC, HigherHRNet, KAPAO	Not stated	Custom 1,400-image dataset from 72 h video	Pose AP 97.1%, AR 97.8%; feeding precision / recall: FB 98.8 / 99.9%; NNVB 99.1 / 91.3%; NFB 100 / 100
102	Liu et al. (2025b)	Hanwoo cattle	Hikvision 4 K cameras	Elevated fixed cameras	TransReID, Pose-guided Heatmap Encoder, VCR	ImageNet-pretrained ViT backbone; pose	HanwooReID, 9,929 cropped	Closed-set mAP up to 95.3%, Rank-1

(*continued on next page*)

(*continued*)

No.	Work	Livestock targeted	Camera technology	Camera location	ML/DL models used	Pretrained networks used	Dataset used	Performance obtained
						estimator pretrained on AP-10 K / Animal-Pose	images from 31 cattle	96.8%; open-set mAP up to 74.8%
103	Wang et al. (2025)	Yaks	Binocular stereo camera	Outdoor field setup	YOLOv8 n/s/m, Keypoint R-CNN baseline, GPR, AdaBoost, GBR, RF, SVR, stacking, voting, ETR	Not stated	7,895 labeled images + 50-yak binocular subset	YOLOv8n keypoint precision 93%; best weight regression with GPR: MAE 25.4 kg, RMSE 31.0 kg, R^2 0.72, MAPE 0.12
104	Xu et al. (2025a)	Chickens	Orbbec Femto Bolt RGB-D cameras	Front + side	CW-Measure-pose, StarNet, FDPN-DASI, C3k2-Star, Wise-ShapeIoU	Not explicitly reported	4,293 images, augmented to 6,800; 1,000 paired RGB-depth test images	Precision 98.66%; recall 98.73%; mAP0.5 99.15%; mAP0.5:0.95 93.30%; average 16.19 s per chicken
105	Rozendo et al. (2025)	Cattle	Public RGB side-view datasets	Side view	CAFE-Net, MMPose ResNet50, Depth Pro, OLS, LASSO, stepwise LR, RF, ANN	CAFE-Net on CattleSegment; MMPose pretrained on AnimalPose; pretrained Depth Pro	Kaggle B3 and B4	Segmentation IoU 0.9738 / 0.9532; best filtered MAPE 11.62% on B3 and 10.02% on B4
106	Reitmaier and Narli (2025)	Holstein-Friesian dairy cows	Intel RealSense D435 RGB	Side view	DeepLabCut with ResNet-152; Random Forest direct and hierarchical models	ResNet-152 backbone, pretraining not stated	424 videos for classification; 7,650 annotated frames	Model-1 accuracy 87.5%; hierarchical model 94%; keypoint test error 6.66 px
107	Li et al. (2025b)	Horses	Orbbec Gemini 2 XL RGB-D	Side-view stationary capture	YOLOv8 detection, segmentation, pose; posture verification rules; contour search; linear regression for CI filtering	Pretrained YOLOv8 models used	3,000 training / validation images + 500 test images; 136 horses for evaluation	IoU > 90%; OKS > 95%; BH MAPE 1.97%; BL 2.56%; CC 3.94%; CBC 7.84%; 35–45 s per horse
108	Pham et al. (2025)	Cattle, beef calves	ICI FMX 400 thermal + RGB webcam	Frontal facial view	Faster R-CNN + FPN with ResNet / ViT / Swin; RF regression best	Standard Detectron2 backbones; explicit pretraining not stated	CattleFace-RGBT, 2,300 paired RGB + thermal images	RGB keypoint AP up to 94.37; thermal AP up to 73.16; fever estimation MSE ~ 0.13–0.21 and R^2 ~ 0.65–0.87
109	Zhao et al. (2025b)	Simmental beef cattle	Intel RealSense D455 RGB-D	Overhead / top-down	DeepLabCut with EfficientNet-B6 backbone, SVM posture filter	Transfer learning in DeepLabCut; explicit details not stated	1,546 image sets from 198 cattle	Relative errors: body length 1.22%, width 2.31%, height 1.53%, hip width 2.50%, hip height 1.80%; ~6 FPS
110	Bibinbe et al. (2025)	Pigs	Single fixed video camera, unspecified	Not reported	YOLOX, DeepLabCut, ByteTrack, DeepSORT, CIC, Track-Anything, PromptTrack, idTracker	SAM in Track-Anything; OWLv2 in PromptTrack	10 min video, 15 pigs, 782 annotated frames	ByteTrack IDF1 0.79, MOTA 0.73; DeepSORT IDF1 0.74; Track-Anything IDF1 0.70
111	Weng et al. (2025)	Cattle	Intel RealSense D455i	Lateral side view	YOLOv8n-pose, YOLOv8-DMC-pose, DRAMiTransformer, MHSA-C2f, CASimAM	Ultralytics / YOLOv8 baseline implied	>7,000 images total including custom RGB-D and public datasets	AP@0.5 0.931; AP@0.50:0.95 0.868; precision 0.882; recall 0.881; 28.3 FPS on Jetson Xavier NX
112	Duan et al. (2025)	Holstein dairy cows	Intel RealSense D455 RGB-D	Overhead	AGC-SimCC with HR-ACRNet, RF, KNN, SVM	HRNet backbone within MMPose; pretraining not specified	741 walking sequences; 2,520 annotated frames	Keypoint AP 95.89%; AR 97.94%; PCK@0.02 100%; lameness RF accuracy 0.91, macro-F1 0.85
113	Zhang et al. (2025b)	Beef cattle	2 K overhead network cameras	Overhead / dorsal	YOLOv11, HRNet, OSNet, ResNeXt50, EfficientNet-B2, MaxViT	Models pretrained on BECA-D / transfer; ImageNet-style details not explicit	BECA-D and BECA-L	Detection precision 97.4%, recall 97.9%, mAP@0.5 99.3%; keypoint mAP@0.5 97.9%; few-shot ReID accuracy up to 93.45%
114	Chen et al. (2025b)	Cattle	Hikvision MV-CU020-19GC RGB industrial cameras	Side + overhead	SCW-YOLO, S2FE, C2DCRA, WTPHead	Not explicitly reported	9,150 images from 61 cattle	~1.2 M params, 2.2 GFLOPs, 2.5 MB; mAP ~ 97.4%; overall body-measurement MRE ~ 5.7%

Data availability

Data will be made available on request.

References

Alsaaod, M., Fadul, M., Steiner, A., 2019. Automatic lameness detection in cattle. Vet. J. 246, 35–44.

Andriluka, M., Pishchulin, L., Gehler, P., Schiele, B., 2014. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3686–3693.

Bai, L., Guo, C., Song, J., 2025. Cattle weight estimation model through readily photos. Eng. Appl. Artif. Intel. 143, 109976.

Barney, S., Dlay, S., Crowe, A., Kyriazakis, I., Leach, M., 2023. Deep learning pose estimation for multi-cattle lameness detection. Sci. Rep. 13 (1), 4499.

Berckmans, D., 2014. Precision livestock farming technologies for welfare management in intensive livestock systems. Revue Scientifique et Technique (international Office of Epizootics) 33 (1), 189–196.

Bibinbe, A.M.S.N., Gagnon, P., Ahloy-Dallaire, J. and Paquet, E.R., 2025. Multi-animal tracking in Transition: Comparative Insights into Established and Emerging Methods. *Smart Agricultural Technology*, p.101543.

Biggs, B., Boyne, O., Charles, J., Fitzgibbon, A. and Cipolla, R., 2020, August. Who left the dogs out? 3d animal reconstruction with expectation maximization in the loop. In *European Conference on Computer Vision* (pp. 195-211). Cham: Springer International Publishing.

Cao, J., Tang, H., Fang, H.S., Shen, X., Lu, C., Tai, Y.W., 2019. Cross-domain adaptation for animal pose estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9498–9507.

Chen, J., Liu, L., Li, P., Yao, W., Shen, M., Ding, Q.A., Liu, L., 2025a. PKL-Track: a Keypoint-Optimized approach for piglet tracking and activity measurement. Comput. Electron. Agric. 237, 110578.

Chen, L., Zhang, L., Tang, J., Tang, C., An, R., Han, R., Zhang, Y., 2024. GRMPose: GCN-based real-time dairy goat pose estimation. Comput. Electron. Agric. 218, 108662.

Chen, X., Guo, X., Li, Y., Liu, C., 2025b. A Lightweight Automatic cattle Body Measurement Method based on Keypoint Detection. Symmetry 17 (11), 1926.

Chen, Z., Huang, Q., Zhang, S., Tian, X., Yin, L., 2025c. Multimodal sow lameness classification method integrating spatiotemporal features. Comput. Electron. Agric. 235, 110363.

Del Pero, L., Ricco, S., Sukthankar, R., Ferrari, V., 2017. Behavior discovery and alignment of articulated object classes from unstructured video. Int. J. Comput. Vis. 121 (2), 303–325.

Deng, H., Yang, G., Xu, X., Hua, Z., Liu, J., Song, H., 2025. Fusion of CREStereo and MobileViT-Pose for rapid measurement of cattle body size. Comput. Electron. Agric. 232, 110103.

Dohmen, R., Catal, C., Liu, Q., 2022. Computer vision-based weight estimation of livestock: a systematic literature review. N. Z. J. Agric. Res. 65 (2–3), 227–247.

Du, A., Guo, H., Lu, J., Su, Y., Ma, Q., Ruchay, A., Marinello, F., Pezzuolo, A., 2022. Automatic livestock body measurement based on keypoint detection with multiple depth cameras. Comput. Electron. Agric. 198, 107059.

Duan, W., Wang, F., Li, H., Liu, N., Fu, X., 2025. Lameness detection in dairy cows from overhead view: high-precision keypoint localization and multi-feature fusion classification. Front. Vet. Sci. 12, 1675181.

Fan, Q., Liu, S., Li, S., Zhao, C., 2023. Bottom-up cattle pose estimation via concise multi-branch network. Comput. Electron. Agric. 211, 107945.

Fang, C., Wu, Z., Zheng, H., Yang, J., Ma, C., Zhang, T., 2024. MCP: multi-chicken pose estimation based on transfer learning. Animals 14 (12), 1774.

Fang, C., Zhang, T., Zheng, H., Huang, J., Cuan, K., 2021. Pose estimation and behavior classification of broiler chickens based on deep neural networks. Comput. Electron. Agric. 180, 105863.

Fang, C., Zheng, H., Yang, J., Deng, H., Zhang, T., 2022. Study on poultry pose estimation based on multi-parts detection. Animals 12 (10), 1322.

Farahnakian, F., Björkman, S., Bloch, V., Pastell, M., Heikkonen, J., 2024. Pose estimation of sow and piglets during free farrowing using deep learning. Journal of Agriculture and Food Research 16, 101067.

Farahnakian, F., Heikkonen, J. and Björkman, S., 2021, December. Multi-pig pose estimation using deeplabcut. In 2021 11th international conference on intelligent control and information processing (ICICIP) (pp. 143-148). IEEE.

Fazzari, E., Romano, D., Falchi, F. and Stefanini, C., 2025. Animal behavior analysis methods using deep learning: A survey. Expert Systems With Applications, p.128330.

Feuser, A.K., Gesell-May, S., Müller, T., May, A., 2022. Artificial intelligence for lameness detection in horses—A preliminary study. Animals 12 (20), 2804.

Fodor, I., Taghavi, M., Ellen, E.D., van der Sluis, M., 2025. Top-view characterization of broiler walking ability and leg health using computer vision. Poult. Sci. 104 (2), 104724.

Fodor, I., van der Sluis, M., Jacobs, M., de Klerk, B., Bouwman, A.C., Ellen, E.D., 2023. Automated pose estimation reveals walking characteristics associated with lameness in broilers. Poult. Sci. 102 (8), 102787.

Gan, H., Li, S., Ou, M., Yang, X., Huang, B., Liu, K., Xue, Y., 2021. Fast and accurate detection of lactating sow nursing behavior with CNN-based optical flow and features. Comput. Electron. Agric. 189, 106384.

Gong, C., Zhang, Y., Wei, Y., Du, X., Su, L., Weng, Z., 2022. Multicow pose estimation based on keypoint extraction. PLoS One 17 (6), e0269259.

Gorssen, W., Winters, C., Meyermans, R., D'Hooge, R., Janssens, S., Buys, N., 2022. Estimating genetics of body dimensions and activity levels in pigs using automated pose estimation. Sci. Rep. 12 (1), 15384.

Guan, Z., Chai, X., 2025. MMVSL: a multi-modal visual semantic learning method for pig pose and action recognition. Comput. Electron. Agric. 229, 109791.

Guo, Q., Pei, D., Sun, Y., Langenhuizen, P.P., Orsini, C.A., Martinsen, K.H., Nordbø, Ø., Bolhuis, L., Bijma, P., 2025. January. Multi-object keypoint detection and pose estimation for pigs. In: 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 466–474.

He, Z., Guo, C., Guo, Z., Lyu, L., Huang, E., Mao, A., Liu, K., 2025. Characterization of social cohesion status of pre-weaning piglets based on lightweight pose estimation. Comput. Electron. Agric. 229, 109716.

Hewitt, C., Mahmoud, M., 2019. In: September. Pose-Informed Face Alignment for Extreme Head Pose Variations in Animals. IEEE, pp. 1–6.

Higaki, S., Matsui, Y., Miwa, M., Yamamura, T., Hojo, T., Yoshioka, K., Vang, A., Negreiro, A., Dórea, J.R., 2024. Leveraging computer vision-based pose estimation technique in dairy cows for objective mobility analysis and scoring system. Comput. Electron. Agric. 217, 108573.

Higaki, S., Menezes, G.L., Ferreira, R.E., Negreiro, A., Cabrera, V.E., Dórea, J.R., 2025. Objective dairy cow mobility analysis and scoring system using computer vision–based keypoint detection technique from top-view 2-dimensional videos. J. Dairy Sci. 108 (4), 3942–3955.

Högberg, N., Berthet, D., Alam, M., Nielsen, P.P., Tamminen, L.M., Fall, N. and Kroese, A., 2025. Exploring pose estimation as a tool for the assessment of brush use patterns in dairy cows. Applied Animal Behaviour Science, p.106746.

Hu, Y., Dai, X., Dai, B., Li, R., Fang, J., Yin, Y., Liu, H., Shen, W., 2025. Feeding behavior recognition of group-housed pigs based on pose estimation and keypoint features discrimination. Comput. Electron. Agric. 239, 111039.

Jia, Z., Zhao, Y., Mu, X., Liu, D., Wang, Z., Yao, J., Yang, X., 2025. Intelligent Deep Learning and Keypoint Tracking-based Detection of Lameness in Dairy Cows. Veterinary Sciences 12 (3), 218.

Jiang, L., Lee, C., Teotia, D., Ostadabbas, S., 2022. Animal pose estimation: a closer look at the state-of-the-art, existing gaps and opportunities. Comput. Vis. Image Underst. 222, 103483.

Kang, X., Zhang, X.D., Liu, G., 2021. A review: development of computer vision-based lameness detection for dairy cows and discussion of the practical applications. Sensors 21 (3), 753.

Keele, S., 2007. *Guidelines for performing systematic literature reviews in software engineering* (Vol. 5). Technical report, ver. 2.3 ebse technical report. ebse.

Khin, M.P., Zin, T.T., Mar, C.C., Tin, P. and Horii, Y., 2022, October. Cattle pose classification system using DeepLabCut and SVM model. In 2022 IEEE 11th Global Conference on Consumer Electronics (GCCE) (pp. 494-495). IEEE.

Kitchenham, B., Brereton, O.P., Budgen, D., Turner, M., Bailey, J., Linkman, S., 2009. Systematic literature reviews in software engineering - a systematic literature review. Inf. Softw. Technol. 51 (1), 7–15.

Klecel, W., Rahael, H., Brooks, S.A., 2025. refineDLC: an advanced post-processing pipeline for DeepLabCut outputs. Biology Methods and Protocols 10 (1), p.bpaf084.

Kresovic, M., Nguyen, T., Ullah, M., Afridi, H., Cheikh, F.A., 2022. In: June. Pigpose: A Realtime Framework for Farm Animal Pose Estimation and Tracking. Springer International Publishing, Cham, pp. 204–215.

Kroese, A., Alam, M., Hernlund, E., Berthet, D., Tamminen, L.M., Fall, N., Högberg, N., 2024a. 3-Dimensional pose estimation to detect posture transition in freestall-housed dairy cows. J. Dairy Sci. 107 (9), 6878–6887.

Kroese, A., Högberg, N., Berthet, D., Tamminen, L.M., Fall, N. and Alam, M., 2024. Exploring the link between cow size and sideways lunging using 3D pose estimation. In 11th European Conference on Precision Livestock Farming, Bologna 9-12 September 2024 (pp. 32-39). European Conference on Precision Livestock Farming.

Kroese, A., Högberg, N., Vicuna, E.D., Berthet, D., Fall, N., Alam, M. and Tamminen, L. M., 2025. Evaluating the automated measurement of abnormal rising and lying down behaviours in dairy cows using 3D pose estimation. Smart agricultural technology, p.101205.

Lawin, F.J., Byström, A., Roepstorff, C., Rhodin, M., Almlöf, M., Silva, M., Andersen, P. H., Kjellström, H., Hernlund, E., 2023. Is markerless more or less? Comparing a smartphone computer vision method for equine lameness assessment to multi-camera motion capture. Animals 13 (3), 390.

Li, C., Yang, Y., Weng, Z., Hernlund, E., Zuffi, S., Kjellström, H., 2024a. Dessie: Disentanglement for Articulated 3D Horse Shape and Pose Estimation from Images. In: Proceedings of the Asian Conference on Computer Vision, pp. 764–783.

Li, J., Liu, Y., Zheng, W., Chen, X., Ma, Y., Guo, L., 2024b. Monitoring cattle ruminating behavior based on an improved keypoint detection model. Animals 14 (12), 1791.

Li, J., Yang, Y., Yao, Y., Zou, H., Guo, X., Xiao, J., Hu, R., Cheng, S., Wang, Y., Wang, Z., Peng, Y., 2025a. A dynamic yak heifer pose estimation model based on keypoint detection for complex environmental monitoring. J. Dairy Sci. 108 (12), 13704–13719.

Li, J., Ye, M., Zhang, X., Li, C., Cao, K., Hu, X., Li, Q., Ma, W. and Wang, Q., 2025. A portable RGB-D based device for unconstrained horse body size measurement for breeding. Smart Agricultural Technology, p.101342.

Li, J.W., Ma, W.H., Li, Q.F., Xue, X.L. and Wang, Z.Q., 2022. Automatic acquisition and target extraction of beef cattle 3D point cloud from complex environment.

Li, R., Wen, Y., Zhang, S., Xu, X., Ma, B., Song, H., 2024c. Automated measurement of beef cattle body size via key point detection and monocular depth estimation. Expert Syst. Appl. 244, 123042.

Li, X., Cai, C., Zhang, R., Ju, L., He, J., 2019. Deep cascaded convolutional models for cattle pose estimation. Comput. Electron. Agric. 164, 104885.

Li, X., Sun, K., Fan, H., He, Z., 2023. Real-time cattle pose estimation based on improved RTMPose. Agriculture 13 (10), 1938.

Li, Z., Zhang, Q., Lv, S., Han, M., Jiang, M., Song, H., 2022. Fusion of RGB, optical flow and skeleton features for the detection of lameness in dairy cows. Biosyst. Eng. 218, 62–77.
Liu, C., Ye, H., Lu, S., Tang, Z., Bai, Z., Diao, L., Wang, L., Li, L., 2023a. Skeleton extraction and pose estimation of piglets using ZS-DLC-PAF. Int. J. Agric. Biol. Eng. 16 (3), 180–193.
Liu, D., Parmiggiani, A., Norton, T., 2025a. MT-SRNet: a Transferable Multi-Task Super-Resolution Network for Pig Keypoint Detection, Segmentation, and Posture Estimation. Comput. Electron. Agric. 237, 110533.
Liu, H., Reibman, A.R., Boerman, J.P., 2020. Video analytic system for detecting cow structure. Comput. Electron. Agric. 178, 105761.
Liu, H., Reibman, A.R., Boerman, J.P., 2023b. Feature extraction using multi-view video analytics for dairy cattle body weight estimation. Smart Agric. Technol. 6, 100359.
Liu, J., Fuentes, A., Han, S., Jeong, Y., Yoon, S., Park, D.S., 2025b. HanwooReID: Multi-view cattle re-identification with pose-aware transformer enhancements. Comput. Electron. Agric. 239, 111117.
Liu, J., Xiao, D., Liu, Y., Huang, Y., 2023c. A pig mass estimation model based on deep learning without constraint. Animals 13 (8), 1376.
Liu, L., Kurban, A., Liu, Y., 2024. Improved YOLOv11pose for Posture Estimation of Xinjiang Bactrian Camels. International Journal of Advanced Computer Science & Applications 15 (12).
Liu, S., Fan, Q., Li, S., Zhao, C., 2023d. MAPoseNet: Animal pose estimation network via multi-scale convolutional attention. J. Vis. Commun. Image Represent. 97, 103989.
Lodkaew, T., Pasupa, K., Loo, C.K., 2023. CowXNet: an automated cow estrus detection system. Expert Syst. Appl. 211, 118550.
Lovarelli, D., Bacenetti, J., Guarino, M., 2020. A review on dairy cattle farming: is precision livestock farming the compromise for an environmental, economic and social sustainable production? J. Clean. Prod. 262, 121409.
Lu, H., Zhang, J., Yuan, X., Lv, J., Zeng, Z., Guo, H., Ruchay, A., 2025. Automatic coarse-to-fine method for cattle body measurement based on improved GCN and 3D parametric model. Comput. Electron. Agric. 231, 110017.
Luo, X., Hu, Y., Gao, Z., Guo, H., Su, Y., 2023. Automated measurement of livestock body based on pose normalisation using statistical shape model. Biosyst. Eng. 227, 36–51.
Lynn, N.C., Kyu, Z.M., Zin, T.T. and Kobayashi, I., 2017, June. Estimating body condition score of cows from images with the newly developed approach. In 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) (pp. 91-94). IEEE.
Ma, W., Sun, Y., Qi, X., Xue, X., Chang, K., Xu, Z., Li, M., Wang, R., Meng, R., Li, Q., 2024. Computer-vision-based sensing technologies for livestock body dimension measurement: a survey. Sensors 24 (5), 1504.
Martin-Cirera, A., Nowak, M., Norton, T., Auer, U., Oczak, M., 2024. Comparison of Transformers with LSTM for classification of the behavioural time budget in horses based on video data. Biosyst. Eng. 242, 154–168.
Menezes, G.L., Alves, A.A., Negreiro, A., Ferreira, R.E., Higaki, S., Casella, E., Rosa, G.J., Dorea, J.R., 2025. Color-independent cattle identification using keypoint detection and Siamese neural networks in closed-and open-set scenarios. J. Dairy Sci.
Narli, S.S., Schmidt, H., Firouzabadi, A., Schönnagel, L., Reich, M.S., Reitmaier, S., 2025. Automated detection of lameness in dairy cattle through computer vision analysis of back shape characteristics. Comput. Biol. Med. 197, 111038.
Nasir, M.F., Fuentes, A., Han, S., Liu, J., Jeong, Y., Yoon, S., Park, D.S., 2025. Multi-camera fusion and bird-eye view location mapping for deep learning-based cattle behavior monitoring. Artif. Intell. Agric. 15 (4), 724–743.
Nasiri, A., Yoder, J., Zhao, Y., Hawkins, S., Prado, M., Gan, H., 2022. Pose estimation-based lameness recognition in broiler using CNN-LSTM network. Comput. Electron. Agric. 197, 106931.
Niknejad, N., Caro, J.L., Bidese-Puhl, R., Bao, Y., Staiger, E.A., 2023. Equine kinematic gait analysis using stereo videography and deep learning: stride length and stance duration estimation. Journal of the ASABE 66 (4), 865–877.
Okour, M., Falque, R., Alempijevic, A., 2024. In: October. Sim2real Cattle Joint Estimation in 3D Point Clouds. IEEE, pp. 1818–1823.
O'Leary, N.W., Byrne, D.T., O'Connor, A.H., Shalloo, L., 2020. Invited review: cattle lameness detection with accelerometers. J. Dairy Sci. 103 (5), 3895–3911.
Ong, K.E., Retta, S., Srinivasan, R., Tan, S. and Liu, J., 2023, December. Cattleeyeview: A multi-task top-down view cattle dataset for smarter precision livestock farming. In 2023 IEEE International Conference on Visual Communications and Image Processing (VCIP) (pp. 1-5). IEEE.
Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., et al., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372, n71.
Peng, C., Cao, S., Li, S., Bai, T., Zhao, Z., Sun, W., 2024. Automated measurement of cattle dimensions using improved keypoint detection combined with unilateral depth imaging. Animals 14 (17), 2453.
Perneel, M., Adriaens, I., Aernouts, B., Verwaeren, J., 2025a. Consistent multi-animal pose estimation in cattle using dynamic Kalman filter based tracking. Smart Agric. Technol., 101014
Perneel, M., Adriaens, I., Verwaeren, J., Aernouts, B., 2025b. Dynamic Multi-Behaviour, Orientation-Invariant Re-Identification of Holstein-Friesian cattle. Sensors 25 (10), 2971.
Pham, T.T., Coffman, E., Kegley, B., Powell, J.G., Zhao, J. and Le, N., 2025. CattleFever: An Automated Cattle Fever Estimation System. *Smart Agricultural Technology*, p.101434.
Qu, H., Leng, A., Deng, Y., Yin, X., Tang, X. and Zhai, S., 2025. A Keypoint Detection-Based Approach for Real-Time Aggressive Behavior Identification in Goats. Applied Animal Behaviour Science, p.106745.
Ramesh, M., Reibman, A.R., 2024. SURABHI: Self-Training using Rectified Annotations-based Hard Instances for Eidetic cattle Recognition. Sensors (basel, Switzerland) 24 (23), 7680.
Ramesh, M., Reibman, A.R. and Boerman, J.P., 2023. Eidetic recognition of cattle using keypoint alignment. Electronic Imaging, 35, pp.279-1.
Reitmaier, S., Narli, S., 2025. Validation of a deep learning model for cattle lameness detection: Comparison of human scorer performance and automated gait analysis. Smart Agric. Technol., 101252
Riekert, M., Klein, A., Adrion, F., Hoffmann, C., Gallmann, E., 2020. Automatically detecting pig position and posture by 2D camera imaging and deep learning. Comput. Electron. Agric. 174, 105391.
Rohan, A., Rafaq, M.S., Hasan, M.J., Asghar, F., Bashir, A.K., Dottorini, T., 2024. Application of deep learning for livestock behaviour recognition: a systematic literature review. Comput. Electron. Agric. 224, 109115.
Romano, E., Brambilla, M., Cutini, M., Giovinazzo, S., Lazzari, A., Calcante, A., Tangorra, F.M., Rossi, P., Motta, A., Bisaglia, C., Bragaglio, A., 2023. Increased cattle feeding precision from automatic feeding systems: considerations on technology spread and farm level perceived advantages in Italy. Animals 13 (21), 3382.
Rozendo, G.B., Dadi, M., Franco, A., Lumini, A., 2025. Cattle weight estimation using 2D side-view images and estimated depth-based 3D modeling. Smart Agric. Technol., 101099
Russello, H., van der Tol, R., Kootstra, G., 2022. T-LEAP: Occlusion-robust pose estimation of walking cows using temporal information. Comput. Electron. Agric. 192, 106559.
Russello, H., van der Tol, R., Holzhauer, M., van Henten, E.J., Kootstra, G., 2024. Video-based automatic lameness detection of dairy cows using pose estimation and multiple locomotion traits. Comput. Electron. Agric. 223, 109040.
Schlageter-Tello, A., Van Hertem, T., Bokkers, E.A., Viazzi, S., Bahr, C., Lokhorst, K., 2018. Performance of human observers and an automatic 3-dimensional computer-vision-based locomotion scoring method to detect lameness and hoof lesions in dairy cows. J. Dairy Sci. 101 (7), 6322–6335.
Shirey, J., Smythe, M.P., Dewberry, L.S., Allen, K., Jain, E. and Brooks, S.A., 2023. DeepLabCut-Display: open-source desktop application for visualizing and analyzing two-dimensional gait parameter data in livestock. *BioRxiv*, pp.2023-10.
Siachos, N., Lennox, M., Anagnostopoulos, A., Griffiths, B.E., Neary, J.M., Smith, R.F., Oikonomou, G., 2024. Development and validation of a fully automated 2-dimensional imaging system generating body condition scores for dairy cows using machine learning. J. Dairy Sci. 107 (4), 2499–2511.
Su, L., Li, M., Zhang, Y., Zong, Z., Gong, C., 2024. Fusion of Target and Keypoint Detection for Automated Measurement of Mongolian Horse Body Measurements. Agriculture 14 (7), 1069.
Taghavi, M., Russello, H., Ouweltjes, W., Kamphuis, C., Adriaens, I., 2024. Cow key point detection in indoor housing conditions with a deep learning model. J. Dairy Sci. 107 (4), 2374–2389.
Tan, Z., Xiao, D., Liu, J., Huang, Y., Liu, Y., 2025. Cascade vision-based weight estimation algorithm for monitoring fattening pig weight. Comput. Electron. Agric. 236, 110384.
Thomsen, P.T., Munksgaard, L., Tøgersen, F.A., 2008. Evaluation of a lameness scoring system for dairy cows. J. Dairy Sci. 91 (1), 119–126.
Van Hertem, T., Tello, A.S., Viazzi, S., Steensels, M., Bahr, C., Romanini, C.E.B., Lokhorst, K., Maltz, E., Halachmi, I., Berckmans, D., 2018. Implementation of an automatic 3D vision monitor for dairy cow locomotion in a commercial farm. Biosyst. Eng. 173, 166–175.
Viazzi, S., Bahr, C., Van Hertem, T., Schlageter-Tello, A., Romanini, C.E.B., Halachmi, I., Lokhorst, C., Berckmans, D., 2014. Comparison of a three-dimensional and two-dimensional camera system for automated measurement of back posture in dairy cows. Comput. Electron. Agric. 100, 139–147.
Volkmann, N., Zelenka, C., Devaraju, A.M., Brünger, J., Stracke, J., Spindler, B., Kemper, N., Koch, R., 2022. Keypoint detection for injury identification during turkey husbandry using neural networks. Sensors 22 (14), 5188.
Wang, F., Zou, X., Chen, Z., Tang, Q., Li, T., Wang, S., Yang, L., Tang, D., 2025. Non-contact weight intelligent estimation based on yak skeleton localization. Eng. Appl. Artif. Intel. 161, 112036.
Wang, X., Wang, W., Lu, J., Wang, H., 2022. HRST: an improved HRNet for detecting joint points of pigs. Sensors 22 (19), 7215.
Wang, Y., Mücher, S., Wang, W., Guo, L., Kooistra, L., 2023a. A review of three-dimensional computer vision used in precision livestock farming for cattle growth management. Comput. Electron. Agric. 206, 107687.
Wang, Z., Zhou, S., Yin, P., Xu, A., Ye, J., 2023b. GANPose: Pose estimation of grouped pigs using a generative adversarial network. Comput. Electron. Agric. 212, 108119.
Wei, Y., et al., 2025. Automatic measurement method of sheep body size based on 3D reconstruction and point cloud segmentation. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2025.110978.
Wei, Y., Zhang, H., Gong, C., Wang, D., Ye, M., Jia, Y., 2023. Study of pose estimation based on spatio-temporal characteristics of cow skeleton. Agriculture 13 (8), 1535.
Weng, Z., Hao, W., Gong, C., Zheng, Z., 2025. YOLOv8-DMC: Enabling Non-Contact 3D cattle Body Measurement via Enhanced Keypoint Detection. Animals 15 (18), 2738.
Weng, Z., Mei, Z., Zheng, Z., Gong, C., 2023. February. Video-based multi-scene cow pose estimation. In: Proceedings of the 2023 3rd International Conference on Bioinformatics and Intelligent Computing, pp. 187–194.
Witte, J.-H., Heseker, P., Probst, J., Traulsen, I., Kemper, N., Marx Gómez, J., 2024. Image-based tail posture monitoring of pigs. In: Proceedings of the 57th Hawaii International Conference on System Sciences, pp. 7831–7839.
Wu, S., Han, S., Mo, X., Wei, Y., Qin, Y., Chen, H., Wu, J., Zeng, Z., 2025. A top-down deep neural network for multi-dairy cows pose estimation and lameness detection. Comput. Electron. Agric. 239, 110911.

Wu, S., Han, S., Zhang, J., Cheng, G., Wang, Y., Zhang, K., Han, M., Wu, J., 2024. Multi-scale keypoints detection and motion features extraction in dairy cows using ResNet101-ASPP network. J. Integr. Agric.

Xu, J., Hu, Y., Gou, Z., Lu, Y. and Cui, D., 2025. Automated non-invasive method for measuring chicken comb and wattle using multi-camera system and CW-Measure-pose. *Expert Systems with Applications*, p.129420.

Xu, Z., Li, Q., Ma, W., Li, M., Morris, D., Ren, Z., Zhao, C., 2025b. A geodesic distance regression-based semantic keypoints detection method for pig point clouds and body size measurement. Comput. Electron. Agric. 234, 110285.

Yang, G., Qiao, Y., Deng, H., Shi, J.Q., Song, H., 2025. One-stage keypoint detection network for end-to-end cow body measurement. Eng. Appl. Artif. Intel. 146, 110333.

Yang, L., Jiang, T., Gui, X., Duan, Q., 2024. Automated body measurement of beef cattle based on keypoint detection and local point cloud clustering. Meas. Sci. Technol. 35 (12), 126013.

Yang, Y., Komatsu, M., Oyama, K. and Ohkawa, T., 2023, June. SCIRNet: Skeleton-based cattle interaction recognition network with inter-body graph and semantic priority. In 2023 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

Yang, Y., Qin, W., Sun, L. and Shi, W., 2022, September. A Pig Pose Estimation Model for Measuring Pig's Body Size. In Computer Graphics International Conference (pp. 28-40). Cham: Springer Nature Switzerland.

Yik, S., Benjamin, M., Lavagnino, M. and Morris, D., 2020, October. DIAT (Depth-Infrared Image Annotation Transfer) for training a depth-based pig-pose detector. In 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 2459-2466). IEEE.

Yu, Z., Wan, L., Yousaf, K., Lin, H., Zhang, J., Jiao, H., Yan, G., Song, Z., Tian, F., 2024. An enhancement algorithm for head characteristics of caged chickens detection based on cyclic consistent migration neural network. Poult. Sci. 103 (6), 103663.

Yue, C., Zhang, X., Zhang, Y., Jiang, H. and Miao, H., 2025. Automatic and Accurate Measurement of Cattle Body Based on a Lightweight YOLOv8-Pose Model and 3D Point Cloud. Machine Learning with Applications, p.100662.

Zhang, H., Ma, Y., Shi, Y., Qiao, Y. and Wang, M., 2024, October. A framework for automatic analysis of sheep ruminating behavior based on keypoint detection. In 2024 30th International Conference on Mechatronics and Machine Vision in Practice (M2VIP) (pp. 1-6). IEEE.

Zhang, Y., Li, L., Li, C., Wang, S., Rong, Y., Niu, K. and He, Z., 2025. BECA: A Computer Vision Dataset for Long-Term Recognition In Beef Cattle. Scientific Data.

Zhang, Y., Li, L., Li, C., Wang, S., Rong, Y., Niu, K. and He, Z., 2025. BECA: A Computer Vision Dataset for Long-Term Recognition In Beef Cattle. *Scientific Data*.

Zhao, E., Chen, B., Zhao, H., Liu, G., Jiang, J., Li, Y., Zhang, J., Luo, C., 2024. Lightweight cattle pose estimation with fusion of reparameterization and an attention mechanism. PLoS One 19 (8), e0306530.

Zhao, K., Zhang, M., Ji, J., Zhang, R., Bewley, J.M., 2023a. Automatic lameness scoring of dairy cows based on the analysis of head-and back-hoof linkage features using machine learning methods. Biosyst. Eng. 230, 424–441.

Zhao, S., Bai, Z., Huo, L., Han, G., Duan, E., Gu, Y., Meng, L., 2025a. Deep learning-based approach for behavior recognition and pose estimation in cage-reared ducks: Leveraging cage-net elimination to propel precision improvement. Comput. Electron. Agric. 236, 110469.

Zhao, S., Bai, Z., Meng, L., Han, G., Duan, E., 2023b. Pose estimation and behavior classification of Jinling white duck based on improved HRNet. Animals 13 (18), 2878.

Zhao, Y., Wang, H., Zhu, J., Wei, B., Guo, H., Tian, M. and Li, B., 2025. Automatic Geometric Parameter Measurement for Beef Cattle Grouping Platform. *Smart Agricultural Technology*, p.101456.

Zin, T.T., Seint, P.T., Tin, P., Horii, Y., Kobayashi, I., 2020. Body condition score estimation based on regression analysis using a 3D camera. Sensors 20 (13), 3705.