



Deep learning based 2D computer vision in precision livestock farming for pigs: A systematic literature review and future research directions

Hassan-Roland Nasser^{a,*}, Claudia Kasper^b, Vladimir Živković^{c,*}

^a Agroscope, Digital Production Group, Tänikon 1, Ettenhausen, 8356, Switzerland

^b Agroscope, Animal PhenoGenomics, Rte de la Tioleyre 4, Posieux, 1725, Switzerland

^c Institute for Animal Husbandry, Department of Pig Breeding, Autoput Beograd-Zagreb, 16, Belgrade, 11000, Serbia

ARTICLE INFO

Keywords:

Smart farming
Deep learning
Precision livestock farming
Pigs

ABSTRACT

The pressing challenges in pig production - particularly in welfare monitoring, behavior analysis, and health management - have driven increasing research efforts in computer vision as a scalable and automated solution. Over the past decade, a diverse body of work has emerged within precision livestock farming (PLF), addressing tasks ranging from pig detection and identification to behavior recognition and health assessment. This expanding research landscape calls for a structured synthesis to consolidate existing knowledge, clarify methodological developments, and elucidate the role of behavior-focused applications. This systematic literature review examines the adoption of computer vision in pig production through four research dimensions: (i) the range of problems addressed, (ii) the models and neural network architectures employed, (iii) the evaluation metrics reported, and (iv) the technical and practical challenges associated with developing and deploying such models. The contribution of this review is threefold. First, it offers an accessible conceptual overview of key deep learning based computer vision approaches to support interdisciplinary researchers entering the field. Second, it synthesizes methodological and evaluative trends across studies. Third, it identifies key challenges and research opportunities for advancing robust and practically relevant computer vision systems in pig farming.

1. Introduction

Recent advances in artificial intelligence (AI), particularly in deep learning (DL), have fundamentally reshaped computer vision (CV) and pattern recognition [1,2]. Modern DL-driven vision systems can automate tasks that previously relied on human observation, enabling continuous monitoring across agricultural production environments and

supporting data-driven decision-making in digital agriculture [3]. Breakthroughs in convolutional neural networks (CNNs) - deep learning architectures specialized for visual data and inspired by the hierarchical structure of the human visual system - have been central to this progress, enabling automated feature extraction through convolutional layers and substantially improving image analysis performance [4].

These methodological advances have driven rapid expansion of CV

List of abbreviations: AI, Artificial intelligence; AP, Average precision; AUC, Area under the curve; BFLOPs, Billion floating-point operations; CASP, Critical Appraisal Skills Programme; CLIP, Contrastive Language-Image Pretraining; CNN, Convolutional neural network; CV, Computer vision; RGB, Red, Green, Blue; DL, Deep learning; DenseNet, Densely Connected Convolutional Network; DETR, Detection Transformer; FPS, Frames per second; FLOPs, Floating-point operations; F1-score, Harmonic mean of precision and recall; GFLOPs, Giga floating-point operations; HOG, Histogram of Oriented Gradients; HOTA, Higher Order Tracking Accuracy; HRNet, High-Resolution Network; IDF1, Identification F1 score; ImageNet, Large-scale image dataset for visual recognition pretraining; IoU, Intersection over Union; LSTM, Long Short-Term Memory; LVLM, Large vision-language model; MAE, Mean absolute error; mAP, Mean average precision; mAP50, Mean average precision at IoU threshold 0.50; mAP50-95, Mean average precision averaged across IoU thresholds from 0.50 to 0.95; mIoU, Mean Intersection over Union; MOTAP, Multiple Object Tracking Accuracy; MOTP, Multiple Object Tracking Precision; MSE, Mean squared error; OpenPose, Open-source real-time multi-person keypoint detection framework; PCK, Percentage of Correct Keypoints; PLF, Precision livestock farming; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RF-DETR, Receptive Field Detection Transformer; RMSE, Root mean square error; RQ, Research question; SfM, Structure from Motion; SIFT, Scale-Invariant Feature Transform; SLAM, Simultaneous Localization and Mapping; SSD, Single Shot Detector; U-Net, U-shaped convolutional neural network architecture; VGG / VGGNet, Visual Geometry Group Network; ViT, Vision Transformer; VLM, Vision-language model; YOLO, You Only Look Once.

* Corresponding authors.

E-mail addresses: hassan-roland.nasser@agroscope.admin.ch (H.-R. Nasser), claudia.kasper@agroscope.admin.ch (C. Kasper), vzivkovic@istocar.bg.ac.rs (V. Živković).

<https://doi.org/10.1016/j.atech.2026.102426>

Received 1 May 2026; Received in revised form 21 July 2026; Accepted 21 July 2026

Available online 22 July 2026

2772-3755/© 2026 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

applications across agriculture and animal husbandry [5–7]. Precision livestock farming (PLF), particularly in pig production, has benefited from this technological shift. The integration of DL-based CV systems enables automated behavioral monitoring and welfare assessment [8], aligning with increasing societal and regulatory expectations regarding animal welfare [9].

Within pig production systems, CV has emerged as a key enabling technology for continuous, automated, and non-invasive assessment under commercial conditions. Early work by [10] demonstrated one of the first reliable systems capable of locating and identifying individual pigs in commercial pens, establishing the foundation for applications requiring accurate individual recognition. Subsequent advances in DL facilitated more sophisticated identification and tracking approaches, including markerless systems capable of distinguishing and following pigs over time [11].

Over the past decade, extensive research has explored CV for high-throughput phenotyping [7], behavior tracking [12], and welfare assessment [13] in farm animals. While various imaging modalities have been explored, most DL-based approaches in pig production rely on 2D image processing.

Several systematic reviews have examined different aspects of CV and DL in livestock systems. For example, [14] reviewed deep-learning approaches for behavior recognition, while [15] analyzed vision-based monitoring techniques for outdoor livestock. Other reviews have emphasized public dataset availability [16], focused on body-weight estimation [17], or examined broader machine-learning applications in livestock [18]. Although these studies provide valuable insights, a pig-focused systematic review of deep learning-based CV applications in precision livestock farming is still lacking.

To address this need, the present review presents three interrelated contributions. First, it provides a structured mapping of DL-based CV applications with examples in pig production, delineating the range of problems addressed within precision livestock farming and offering an accessible entry point for researchers seeking to engage with this domain. Second, it systematically maps the landscape of deep learning-based computer vision applications in pig production, synthesizes methodological and evaluative practices across studies, and critically examines the technical and practical challenges that shape their development and deployment by answering four research questions. Finally, by consolidating these dimensions, the review outlines future directions in precision livestock farming.

2. An introduction to computer vision with examples in livestock for pigs

The objective of this section is twofold. First, it serves as an accessible introduction to CV concepts for researchers who may not have a technical background in the field but are interested in understanding its potential applications in PLF for pigs. Second, it delineates the specific subfields of CV that are most relevant to this review, helping to accurately clarify the scope and focus given the breadth of the CV domain.

We define CV broadly as a spectrum of computational techniques aimed at enabling machines to interpret and extract information from visual inputs. For the purposes of this review, we structure the field into five principal areas, each representing a step forward in technical sophistication and automation capabilities:

1. **Image processing:** This foundational layer operates at the pixel or signal level and applies deterministic algorithms to enhance or transform images without performing semantic inference. Typical operations include noise reduction, filtering, contrast enhancement, thresholding, morphological transformations, and background subtraction. The resulting images may improve visual quality or isolate regions of interest and can either be used directly or serve as inputs to higher-level computer vision models.

2. **Traditional (classical) computer vision:** Classical computer vision builds upon image processing by extracting structured representations from images using handcrafted features and explicit mathematical models. Techniques include edge detection, feature extraction (e.g., SIFT, HOG) [19], optical flow, template matching, and classical object tracking algorithms. These engineered descriptors are typically combined with conventional machine-learning methods such as support vector machines or decision trees for tasks like detection, classification, or motion analysis. While effective in constrained settings, these approaches require task-specific feature design and substantial parameter tuning.

3. **Geometric computer vision:** Geometric vision focuses on recovering spatial structure and motion from images using principles of projective geometry and multi-view consistency. Because a single monocular image does not contain explicit metric depth information, geometric approaches often rely on stereo vision, multi-camera systems, or sequential image acquisition to infer three-dimensional structure. Representative methods include structure-from-motion (SfM), simultaneous localization and mapping (SLAM), bundle adjustment, and multi-view reconstruction. These techniques are widely used for 3D mapping, pose estimation, and spatial modeling in robotics and remote sensing.

4. **Deep learning-based computer vision:** Deep learning has transformed computer vision by enabling automatic learning of hierarchical feature representations directly from data. CNNs initially drove this progress by learning spatial feature hierarchies, where early layers capture low-level patterns (e.g., edges or textures) and deeper layers encode increasingly abstract concepts (e.g., objects or behaviors). More recently, transformer-based architectures and hybrid models have further advanced visual recognition tasks. Pre-trained backbone networks can be fine-tuned for applications such as object detection, semantic segmentation, pose estimation, and classification. Unlike classical approaches, feature representations are learned rather than manually engineered, improving generalization across diverse environments.

5. **Vision-language models and foundation models:** Vision-language models (VLMs) extend deep learning-based vision systems by jointly modeling visual and textual data within large-scale pretrained architectures. Trained on massive multimodal datasets, these models can perform cross-modal reasoning and exhibit zero-shot or few-shot capabilities for certain tasks. Examples include contrastively trained models (e.g., CLIP-like architectures) and generative multimodal transformers. Although still emerging in livestock applications, such models hold potential for integrating visual observations (e.g., animal behavior) with textual metadata (e.g., health records or environmental logs) to enable multimodal decision support systems.

Given the rapid advancement and dominant role of deep learning in contemporary computer vision, the remainder of this review focuses specifically on DL-based 2D approaches and their application in pig production systems. In this review, *2D computer vision* refers specifically to methods operating on two-dimensional images and video frames acquired with standard visible-light (RGB) cameras. The scope therefore includes RGB still images and RGB video, and excludes thermal or infrared imaging, depth-derived and 3D representations (e.g., RGB-D, stereo, and point-cloud data), and other non-RGB or multimodal sensing. These modalities were excluded because they entail different annotation requirements, model architectures, and on-farm deployment constraints, and because RGB cameras remain the most widely available and readily deployable sensing modality in commercial pig production. Consistent with this scope, studies relying on thermal or 3D imagery were removed at the critical-appraisal stage. In practical implementations, DL-based vision pipelines typically follow a modular architecture composed of a pretrained feature extractor (commonly referred to as a *backbone*) coupled with task-specific prediction heads that perform

inference on well-defined downstream objectives. This architectural decomposition provides a structured lens through which existing methods in PLF can be analyzed.

Pretrained backbone networks function as hierarchical feature extractors, progressively transforming raw pixel data into increasingly abstract representations. Widely adopted CNN architectures include Visual Geometry Group Networks (VGGNet) [20], Residual Networks (ResNet) [21], EfficientNet [22], and Densely Connected Convolutional Networks (DenseNet, [23]). Many of these models are pretrained on large-scale datasets such as ImageNet [24], enabling transfer learning in agricultural settings where annotated data are often limited. More recently, transformer-based architectures such as Vision Transformers (ViTs) [25], as well as hybrid CNN-transformer models, have gained traction due to their ability to capture long-range spatial dependencies and contextual relationships; an advantage in crowded and visually complex environments such as commercial pig pens.

Depending on the research objective, these backbone networks are combined with specialized prediction modules, giving rise to several canonical downstream tasks in computer vision:

- **Classification** (Fig. 1): Classification assigns a single label to an entire image or short image sequence without explicit object localization. In pig production systems, classification models are commonly used to recognize postures, activities, or health-related states. Typical applications include posture recognition (e.g., lateral or ventral lying) and behavior categorization. CNN-based architectures such as ResNet [21], VGG [26], and EfficientNet [22] are frequently employed. For short video clips, 3D CNNs [27] extend this framework by jointly modeling spatial and temporal information.
- **Object Detection** (Fig. 2 (a-b)): Object detection simultaneously localizes and classifies objects within an image, typically through bounding box prediction and associated class labels. In livestock monitoring, detection models are used to identify individual pigs, count animals in pens, localize body parts, or detect specific interactions. Common architectures include two-stage detectors such as Faster R-CNN [28] and one-stage detectors such as YOLO (You Only Look Once, [29]) and SSD (Single Shot Detector, [30]). Recently, another architecture, the transformer-based [31] detectors, including DETR [32] and its variants such as RF-DETR [33],

have gained attention due to their end-to-end formulation and improved handling of complex scenes.

- **Semantic and Instance Segmentation** (Fig. 2 (c)): Segmentation predicts a class label for **each pixel**, rather than assigning a single label to the entire image or to a bounded region. In semantic segmentation, pixels are categorized by class (e.g., animal versus background), whereas instance segmentation additionally distinguishes between individual objects, enabling separation of individual pigs within a group. Pixel-level delineation provides spatially explicit representations that support quantification of posture, locomotion, social interactions, and spatial distribution within pens, as well as extraction of specific body regions such as head, tail, or ears. Representative architectures include U-Net [34], DeepLabV3+ [35], and Mask R-CNN [36] which integrates detection and instance segmentation.
- **Pose Estimation** (Fig. 2 (d)): Pose estimation detects anatomical keypoints (e.g., ears, nose, neck, vertebral column, tail, legs) and infers skeletal structure, enabling fine-grained characterization of body configuration and movement. In pig production, pose estimation has been used to monitor locomotion, detect lameness, and assess recovery after injury. Temporal analysis of keypoint trajectories allows automated quantification of behaviors such as standing, walking, or lying. State-of-the-art frameworks include OpenPose [37], HRNet [38], and DeepLabCut [39] which have been adapted to animal pose estimation tasks.
- **Action Recognition** (Fig. 3): Action recognition extends image-based classification to video sequences by explicitly modeling temporal dynamics. The objective is to identify complex behaviors that unfold over time, such as aggression, social interaction, or farrowing events. In group-housed pig systems, one study combined 3D CNNs with a temporal attention mechanism to detect aggressive interactions and distinguish play from conflict behaviors with high precision [40]. Common architectures for temporal modeling include 3D CNNs, ConvLSTM networks [41], and transformer-based sequence models.

3. Methodology

3.1. Review protocol

This systematic review followed the PRISMA 2020 guidelines [42].

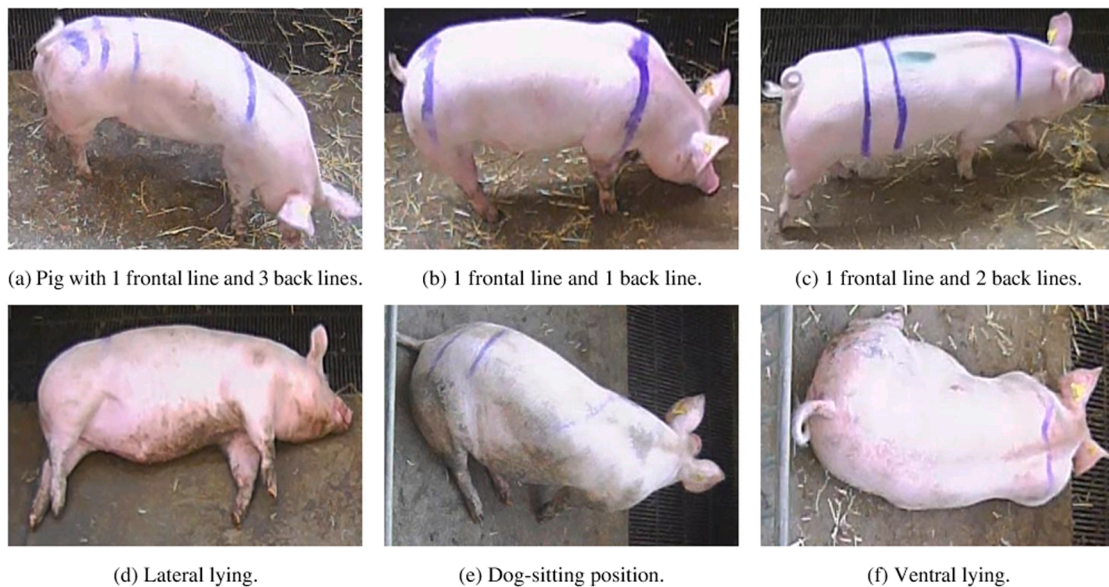


Fig. 1. Example of image classification. (a)–(c) show pigs classified according to the number and position of spray lines on their bodies. (d)–(f) show pigs classified according to their pose.

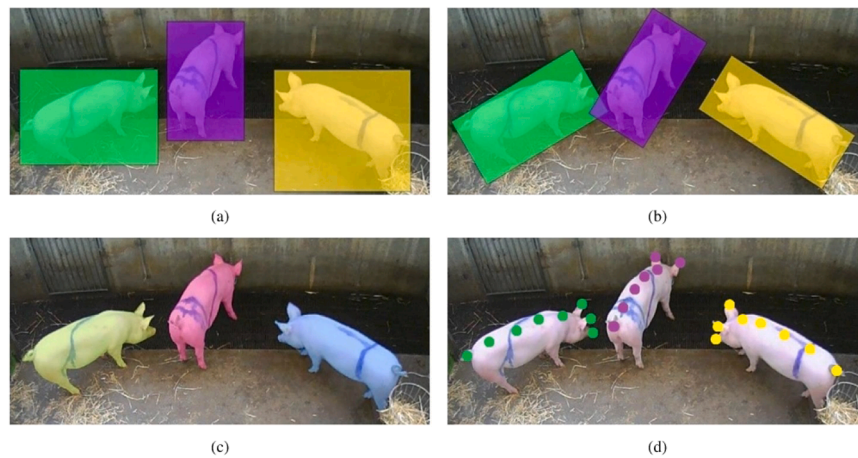


Fig. 2. Examples of multiple computer vision downstream tasks. (a) object detection with bounding boxes; (b) object detection with oriented bounding boxes; (c) semantic segmentation and (d) keypoints detection.

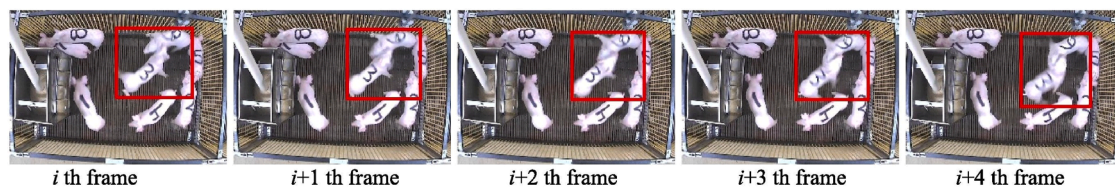


Fig. 3. Example of a frame sequence used for behavior recognition using CNN and LSTM (Long Short-Term Memory) models. Spatial features are extracted from individual frames by the CNN, while the LSTM models temporal dependencies across frames to capture the evolution of the action over time. Reproduced from [39].

The protocol comprised four stages: planning, execution, quality assessment and data extraction, and reporting.

Step 1: Planning. The planning phase defined the review rationale, research questions, search strategy, and explicit inclusion and exclusion criteria. Relevant scientific databases were identified, and search strings were constructed to capture DL-based computer vision studies in pig-focused precision livestock farming.

Step 2: Execution. Search queries were applied to titles, abstracts, and keywords in the selected databases. Retrieved records were screened against the predefined eligibility criteria, followed by full-text assessment. Backward and forward citation tracking was conducted to identify additional relevant studies.

Step 3: Quality assessment and data extraction. Eligible studies were appraised using predefined quality criteria. For each included paper, data were extracted regarding study objectives, PLF application category, datasets, computer vision models and architectures, evaluation protocols, publication metadata, and reported technical challenges.

Step 4: Reporting. Findings were synthesized according to the four research questions. Results were presented through narrative analysis and supported by tables and figures summarizing model usage, datasets, application domains, evaluation metrics, and other characteristics across the included studies.

3.2. Research questions

Four research questions (RQs) were formulated to guide data extraction and synthesis:

- RQ.1: What are the problems solved using CV for PLF in pig production?
- RQ.2: What are the CV models and networks used for PLF in pig production?

- RQ.3: What are the evaluation metrics used for CV models for PLF in pig production?
- RQ.4: What are the challenges associated with the CV-based pig identification, detection, behavior and health monitoring?

3.3. Databases and search strategy

The following three databases were selected for this study: ScienceDirect, Scopus, and Web of Science. The initial search process began with the formulation of keyword-based search strings. At the outset, the keywords “pig” OR “sow” OR “weaner” OR “piglet” OR “fatten*” were used to retrieve all relevant studies. However, the results included studies unrelated to the focus area, such as those pertaining to human research. To refine the search and ensure a focus on computer vision applications, additional keywords such as “computer vision” OR “machine vision” were incorporated.

A simplified search strategy was deliberately employed by avoiding overly specific terms. For instance, words such as “action” and “activity” (related to “behavior”), “artificial intelligence” (related to “deep learning”), and “analysis,” “detection,” and “classification” (related to “recognition”) were not explicitly included in the search strings. This approach allowed for a broader retrieval of relevant studies, even when such terms were absent from the metadata of the selected databases.

The following database-specific search queries were used:

- **Scopus:** TITLE-ABS-KEY(("pig" OR "sow" OR "weaner" OR "piglet" OR "fatten*" OR "swine") AND ("computer vision" OR "machine vision"))
- **ScienceDirect:** Title, Abstract, Keywords: (("pig" OR "sow" OR "weaner" OR "piglet" OR "fatten*" OR "swine") AND ("computer vision" OR "machine vision"))
- **Web of Science:** ALL= (("pig" OR "sow" OR "weaner" OR "piglet" OR "fatten*" OR "swine") AND ("computer vision" OR "machine vision"))

The literature search was performed in June 2025, and the review

includes studies published up to this date. No additional search updates were conducted prior to submission.

3.4. Selection criteria

The initial search yielded a large number of irrelevant publications. To refine the selection and ensure alignment with the research objectives, inclusion and exclusion criteria were established to identify publications relevant to the research questions [43]. Each publication was systematically evaluated against these criteria. A publication was selected if all exclusion criteria were false and all inclusion criteria were true.

Following the database queries, 1068 studies were initially retrieved. After removing duplicates, 674 papers were selected for metadata screening (Fig. 4). At this stage, publications were excluded if:

- The publication was not related to pigs or did not employ deep learning computer or machine vision techniques.
- The publication was not written in English.
- The full text was unavailable.
- The publication was a book chapter, conference abstract, data article, mini-review, short communication, thesis, review, or survey article.
- The publication abstract was missing.
- The publication was a preprint or had not undergone peer review.
- The publication was published before 2015.

After applying these criteria, 162 articles (Fig. 4) were advanced to full-text evaluation using the CASP-based checklist [44]. All titles and abstracts of the retrieved records were then screened by V. Živković and H.-R. Nasser. The full texts of the remaining records were independently screened and discussed by V. Živković and H.-R. Nasser. To find further articles covering the topic, V. Živković screened the reference list of the shortlisted articles and the articles that cited the shortlisted articles. The results were again discussed and selected by two reviewers (V. Živković and H.-R. Nasser). Finally, data extraction was performed independently by two authors (V. Živković and H.-R. Nasser). The extracted data were subsequently cross-checked, and any disagreements were resolved through discussion and consensus. Because all disagreements were resolved by consensus, no formal inter-reviewer agreement was calculated.

3.5. Critical appraisal

The critical assessment relied on ten criteria that included relevance, model clarity, application specificity, study design transparency, data set detail, evaluation quality, reproducibility, presentation of the results, discussion of limitations, and contribution to research gaps (Table 1).

Each of these criteria were checked to rate the articles. Per criteria met, half or full point was given. Thus, a maximum of 10 points could be scored. Detailed scoring can be found in the supplementary material. During the critical appraisal step, 31 papers were further excluded because they were not related to the topic (thermal or 3D imagery, not

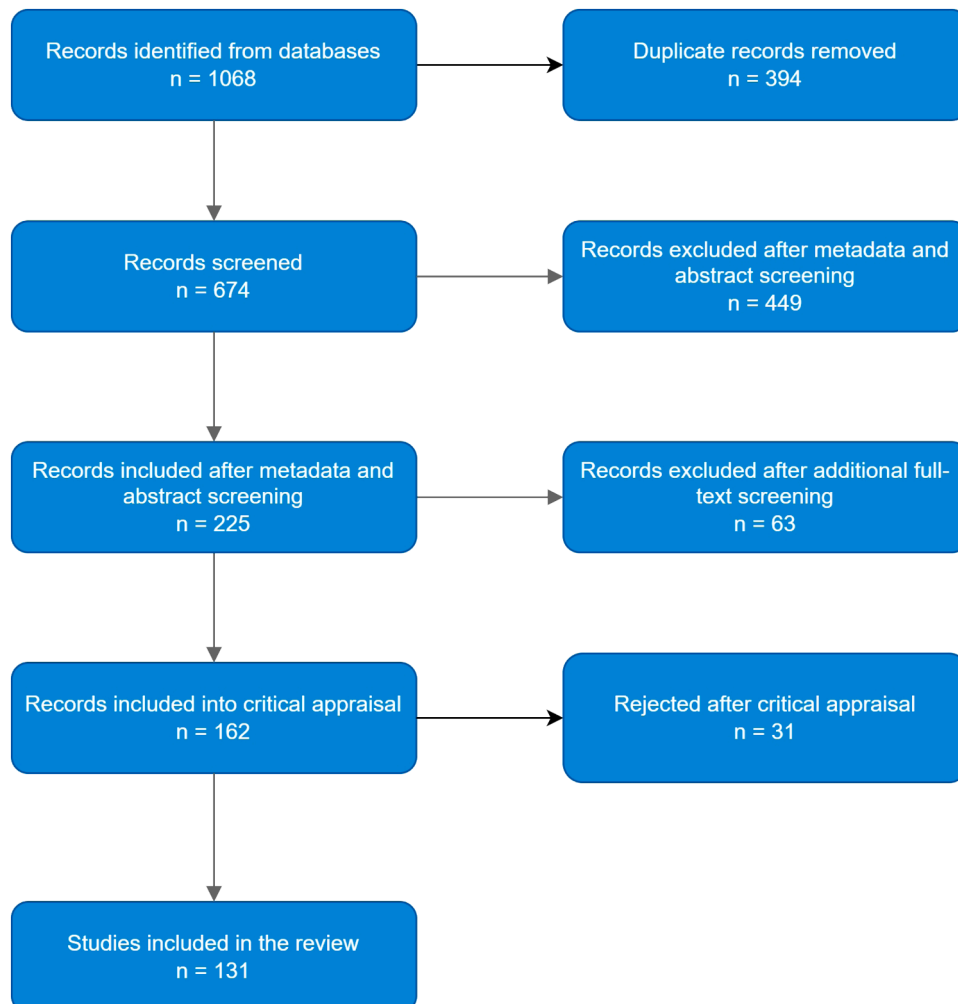


Fig. 4. PRISMA flow diagram illustrating the study selection process for this review.

Table 1

Criteria used to assess studies included in the review, with corresponding descriptions and maximum scores (total possible score = 10).

No.	Criterion	Description	Max score
1	Relevance to review scope	Focus on pigs and use of computer vision techniques	1 (0.5 + 0.5)
2	Computer vision pipeline clarity	CV task clearly defined and model or preprocessing steps explained	1 (0.5 + 0.5)
3	Application area defined	Application area clearly linked to pig welfare or productivity	1
4	Study design transparency	Experimental setup described and rationale provided	1 (0.5 + 0.5)
5	Evaluation quality	Performance metrics reported and CV-relevant evaluation methods used	1 (0.5 + 0.5)
6	Dataset description	Sufficient details on dataset characteristics and annotations provided	1 (0.5 + 0.5)
7	Reproducibility (code/software)	Source code shared or software tools clearly stated	1 (0.5 + 0.5)
8	Results detail	Quantitative results reported with visualizations or examples	1 (0.5 + 0.5)
9	Bias and limitations discussed	Discussion of weaknesses such as occlusions, farm variability, or biases	1 (0.5 + 0.5)
10	Contribution to research gaps	Study offers new insights, roadmap, or highlights missing elements	1

related to deep learning) and therefore were not included in the final list for the critical appraisal. A distribution of appraisal scores is shown in (Fig. 5).

3.6. Data extraction and synthesis

The 131 primary studies (scoring ≥ 8 on the appraisal checklist) were systematically analyzed using a structured extraction framework aligned with the four research questions. A standardized spreadsheet (provided in the supplementary material) was developed to ensure consistent data collection across studies. Each row represented a study, and variables were organized according to the corresponding research question.

Table 2 summarizes the categories of information extracted for each research question.

3.7. Overview of the included studies

Before addressing the research questions in detail and to characterize

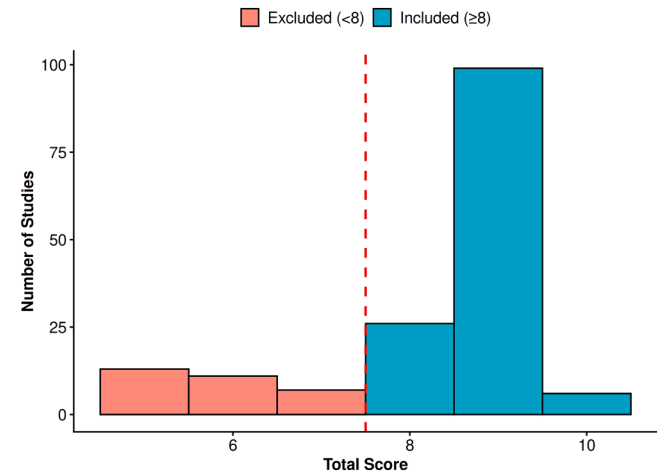


Fig. 5. The distribution of critical appraisal scores. The red dotted line shows the cutoff score for inclusion in the next stage; orange bars reflect papers scoring less than 8, and blue bars those scoring 8 or higher.

Table 2

Categories of information (extracted variables) collected for each research question during data extraction.

Research Question	Extracted Variables
RQ1: Problems addressed	Animal category; application domain; Visiotype; PLF application type; subapplication; behavioral focus; study objectives.
RQ2: Models and architectures	Computer vision task (classification, detection, segmentation, pose, action recognition); CNN architecture name; model family; transformer model (if applicable); tracking algorithms; downstream modeling approaches.
RQ3: Evaluation methods and performance metrics	Evaluation metrics including Accuracy, Precision, Recall, Specificity, Sensitivity, AUC, F1-score, F-measure, mAP, mAP50, mAP50-95, IoU, mIoU, MAE, MSE, RMSE, PCK, R ² , Correlation coefficient, FPS, FLOPs/GFLOPs/BFLOPs, HOTA, MOTA, IDF1, and other task-specific metrics.
RQ4: Technical and practical challenges	Reported challenges such as occlusion, overlapping animals, behavioral complexity, lighting variability, camera position, image quality, posture recognition difficulty, hardware limitations, pen environment constraints, dataset availability, annotation tools, labeled data volume, image resolution, and code availability.
Additional information	Publication year; journal; authorship; dataset characteristics (number of labeled images/videos, training/testing split); image size; software and code availability.

the current research situation on the topic, we provide a descriptive overview of the metadata of the selected studies, including the journals in which they were published, the most active authors, and the distribution of pig categories represented. Fig. 6 summarizes the journals that published at least three of the studies included in this review. Computers and Electronics in Agriculture is the most prominent journal, with 28 papers meeting the inclusion criteria. It is followed by Biosystems Engineering with 9 papers and Agriculture with 7 papers. Several other journals contribute a smaller number of articles, indicating that research on CV-based PLF in pigs is concentrated in a limited set of specialized venues but also present in a broader range of agricultural and engineering journals.

Fig. 7 depicts the distribution of publications by author among the selected studies. T. Norton appears as the most prolific contributor, with 14 papers included in this review, followed by Y. Xue and M. Oczak, each with 7 publications. A long tail of authors is represented by one or a few papers, reflecting a mix of established research groups and emerging contributors to the field.

The distribution of pig categories across the selected studies is shown

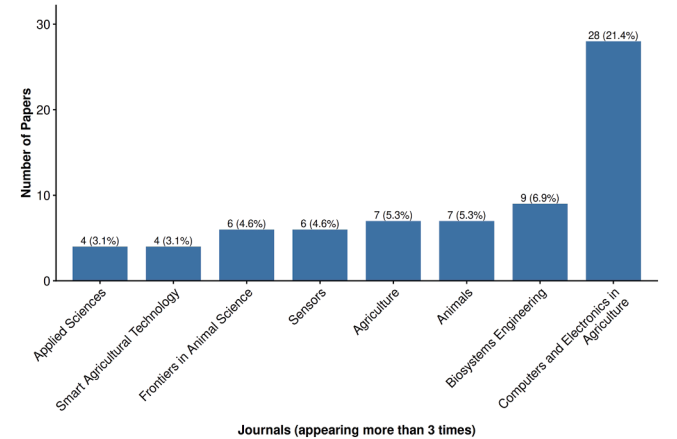


Fig. 6. Distribution of the top selected papers (N = 131) published per journal (only journals with at least three included papers are shown).

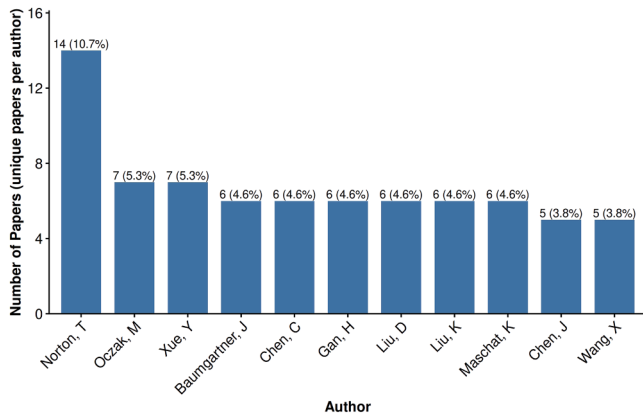


Fig. 7. Distribution of the papers included in the study by the most active authors (top 25).

in Fig. 8. Adult pigs are markedly over-represented, whereas piglets and sows appear less frequently as focal subjects. This imbalance suggests that current CV research in PLF has largely emphasized growing-finish pigs, with comparatively fewer efforts dedicated to early life stages or reproductive animals.

4. Definitions

To ensure terminological clarity in the subsequent sections, we define here the four key terms used to structure and populate the extracted dataset: application, visiotype, behavior, and behavioral category. The extracted dataset itself is provided as an Excel file in the supplementary materials.

Application refers to the production-level objective for which a sensing or computer vision system is deployed within a precision livestock farming framework. It defines the management purpose of the technology, such as health monitoring, welfare assessment, reproduction management, breeding, productivity assessment, or activity monitoring. In other words, the application specifies why the system is implemented in the production context. The applications represented in this dataset are: *Activity Monitoring, Welfare, Breeding and Reproduction, Productivity Assessment, Health Status*. **Visiotype** The term visiotype is used here to denote any algorithmically extracted, biologically or operationally interpretable visual variable derived from animal imagery, spanning a hierarchy from perception-level proxies (e.g., detection, tracking, counting) to higher-level inferences (e.g., behavior recognition, weight estimation). In precision livestock farming (PLF), visiotypes correspond to algorithmically extracted outputs that serve as observable

proxies for an animal's biological state, condition, or behavior and that support monitoring and decision-making.

Visiotypes are not restricted to any single level of the computer vision processing hierarchy. Rather, they span a continuum from perception-level proxies (e.g., detection, tracking, counting) to higher-level inferences interpreted within a biological or production context (e.g., behavior recognition, weight estimation, estrus detection). What unites them is not their algorithmic complexity but their interpretability as management-relevant variables: a low-level output such as detection is itself a visiotype when it answers a production question (for instance, counting animals or estimating space occupancy) rather than serving only as an intermediate step toward a downstream task. In this sense, visiotypes capture what the system ultimately quantifies, detects, estimates, or tracks in relation to the animal and its environment, and the layered relationships among them are described further in Section 7.

The visiotypes included in this dataset are: *Behavior Recognition, Body Part Detection, Body Size Estimation, Counting, Detection, Estrus Detection, Face Recognition, Farrowing Behavior, Feeding Behavior, Identification, Occlusion Resistant Piglets Detection, Occlusion Robust Piglets Detection, Posture Recognition, Re-Identification, Respiration Rate Monitoring, Shoulder Lesion Detection, Stress Monitoring, Tracking, and Weight Estimation*.

Behavior denotes a specific, observable, and operationally defined action that can be directly annotated or detected by a computational model. In this review, the term behavior is used as an umbrella concept that encompasses multiple ethological domains, including social, feeding, exploratory, reproductive, maternal, excretory, damaging, and postural behaviors. Each behavior therefore represents the most granular level of behavioral description in the dataset and belongs to one of these broader behavioral categories. For instance, actions such as fighting or head-to-body knocking are classified as aggressive behavior, tail or ear biting fall under damaging behavior, feeding and drinking belong to feeding behavior, while actions such as standing, sitting, or lying correspond to postural behavior. The behaviors included in this dataset are: *Fighting, Tail Biting, Head-to-body Knocking, Mounting, Ear Biting, Head-to-head Knocking, Body Pressing, Head Biting, Urination, Engagement with Enrichment Objects, Investigating, Rooting, Drinking, Eating, Nursing, Suckling, Lactating, Nest Building, Standing, Lying Down, Sitting, Lateral Lying, Sternal Lying, Moving, Kneeling, Farrowing, Playing, and Nosing*.

Behavioral category refers to a broad ethological domain describing a biologically meaningful class of actions expressed by pigs. It captures functional activity systems that group related actions according to their biological role within the production environment. The behavioral domains represented in this dataset are: *Aggressive Behavior, Excretory Behavior, Exploratory Behavior, Damaging Behavior, Feeding Behavior, Maternal Behavior, Postural Behavior, Reproductive Behavior, and Social Behavior*.

5. Results

This section synthesizes the findings of the 131 included studies in alignment with the four research questions. Quantitative summaries (e.g., distributions, heatmaps, and temporal trends) are complemented by qualitative interpretation to identify dominant application domains, prevailing modeling strategies, evaluation practices, and persistent technical challenges. A consolidated, task-based overview integrating these four dimensions for each task category is provided in Table 3 (Section 5.7).

5.1. RQ1: What are the problems solved using CV for PLF in pig production?

To address RQ1, we identified primary visiotypes (defined in Section 4). A summary of these associations is illustrated in Fig. 9, which uses a heatmap to reveal frequently co-occurring pairs. Additionally, Table 4 shows a detailed listing of the mapping between Applications, Visiotypes

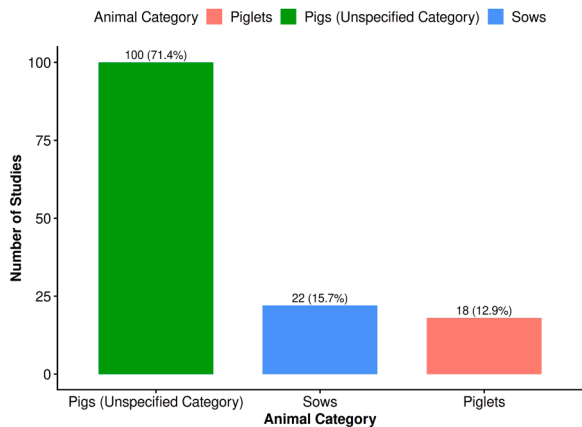


Fig. 8. Distribution of papers (N=131) by pig category (pigs (unspecified category), sows, piglets).

Table 3

Task-based (visiotype) synthesis of the reviewed deep-learning 2D computer-vision studies in pig PLF.

Visiotype (task category)	n	Input	Annotation type	Common model families	Typical evaluation metrics	Main reported limitation
Behavior recognition	41	RGB video / image	Varies: boxes, keypoints, or clip/frame labels	YOLO, ResNet/VGG/EfficientNet, 3D-CNN, + LSTM/ConvLSTM, ST-GCN	Accuracy, precision, recall, F1	Behavioral complexity/ambiguity; temporal modeling; rare-class imbalance
Posture recognition	26	RGB image	Varies: boxes or class labels (some keypoints/masks)	YOLO, Faster R-CNN, ResNet, Mask R-CNN, HRNet	Accuracy, precision, recall, F1, mAP	Posture ambiguity; occlusion; imbalance of rare postures
Detection	21	RGB image	Bounding boxes	YOLO (v3–v10), Faster R-CNN, SSD, RetinaNet, DETR/RF-DETR	mAP, mAP50, mAP50–95, AP50/AP75, precision, recall	Occlusion and overlap in group housing; variable illumination
Tracking	16	RGB video	Boxes + consistent track IDs	Detector (YOLO) + SORT/DeepSORT/ByteTrack/BoT-SORT; FairMOT	MOTA, MOTP, IDF1, HOTA, ID-switches	Identity switches under occlusion/crowding; long-term re-association
Identification / Face / Re-ID	12	RGB image / video	Individual ID labels (crops, faces); varies	CNN classifiers (ResNet, DenseNet, ShuffleNet, MobileNet), ViT, Re-ID networks	Accuracy / top-1, F1 (Re-ID: rank-1, mAP)	Appearance change over time; visually similar individuals; re-ID drift
Counting	8	RGB image	Varies: points/dots, boxes, or density maps	Density networks (CSRNet, MCNN, CClustNet), YOLO, segmentation	MAE, RMSE, counting accuracy	Dense crowding and occlusion; scale variation
Weight / Body-size estimation	8	RGB image (top/side view)	Masks / keypoints / morphometric landmarks	Mask R-CNN / segmentation + regression, EfficientNet, CNN regressors	RMSE, MAE, MAPE, R ² , correlation	Calibration / scale reference; posture and viewpoint variation
Other specialized visiotypes	14	RGB image / video	Task-dependent (boxes, masks, keypoints, or labels)	Task-specific (YOLO, Faster/Mask R-CNN, RetinaNet, CenterNet)	Task-dependent (mAP, accuracy, RMSE)	Small per-task samples; occlusion; subtle or rare clinical/reproductive signs

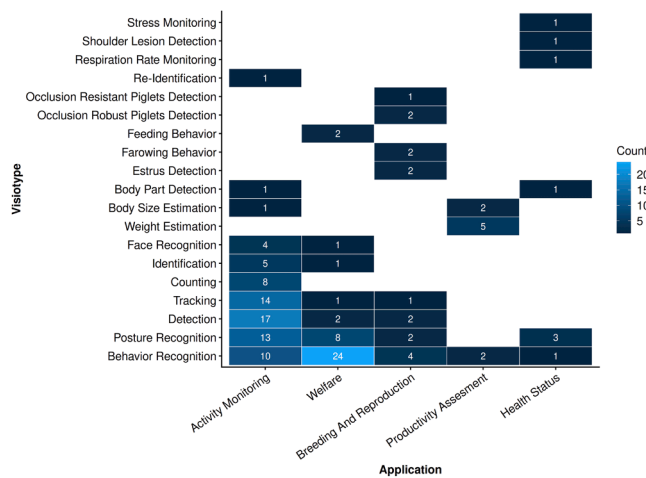


Fig. 9. Heatmap of visiotypes vs applications (distributed by the number of studies). Eight applications are shown on the x-axis and 19 visiotypes on the y-axis. The intensity of the color reflects the number of times these combinations appeared in the screened articles. The most common combination is Welfare using Behavior Recognition as a visiotype, followed by Activity Monitoring using Detection and Tracking. Some combinations are not shown, either because they are not possible or because they have not yet been explored. For example, Health Status never appeared using Detection or Tracking as visiotype.

with a references to each study. Further, it lists additional information needed for further sections.

Activity monitoring represents the predominant application of computer vision (CV) in precision livestock farming (PLF) for pig production ($n = 74$). Across these studies, activity monitoring is inferred using ten distinct Visiotypes. The most frequently employed within this category are Detection ($n = 17$), Tracking ($n = 14$), Posture Recognition ($n = 13$), and Behavior Recognition ($n = 10$). Methodologically, activity monitoring predominantly relies on movement-based visual indicators. Detection and tracking constitute the foundational components for assessing locomotion and general activity dynamics [11,98,171]. Detection is slightly more prevalent, as it is sufficient for frame-level tasks such as animal counting and estimating space occupancy [64].

In contrast, tracking enables the temporal association of individuals across frames and is therefore particularly relevant for quantifying locomotion trajectories and movement dynamics [102]. Behavior recognition typically builds upon tracking outputs, leveraging spatio-temporal movement information to quantify locomotor activity, social interactions, and broader behavioral patterns, which constitute key objectives of activity monitoring systems [172,173]. Posture recognition, by comparison, is used to discriminate between standing, lying, and sitting positions in order to characterize activity states [88,174]. The selection between behavior recognition and posture recognition appears to depend on the granularity of analysis required and the specific biological question under investigation.

The second most represented application is Welfare ($n = 39$), with 7 Visiotypes, driven primarily by Behavior Recognition ($n = 24$) and Posture Recognition ($n = 8$). In this context, behavior recognition was frequently employed to detect abnormal or harmful behaviors such as tail-biting and aggression [138,147], while posture recognition was used to assess inactivity, discomfort, or other welfare-related indicators [163,175]. Behavior and posture based Visiotypes are particularly suitable for welfare assessment, as many welfare impairments manifest through observable changes in activity patterns and social interactions.

Breeding and Reproduction ($n = 16$) shows 8 Visiotypes distributions. Alongside Behavior Recognition ($n = 4$), these application rely mainly on Estrus Detection ($n = 2$) and Farrowing Behavior ($n = 2$), with Occlusion-Robust Piglet Detection ($n = 2$) and Occlusion-Resistant Piglet Detection ($n = 1$). Estrus detection models typically identified mounting behavior and increased activity levels [115], while farrowing behavior systems monitored nest-building activity and posture transitions prior to partus [117]. Piglet detection approaches were particularly relevant for monitoring sow–piglet interactions and reducing piglet crushing risk [118]. Despite their direct relevance to reproductive efficiency and piglet survival, these applications remain comparatively underrepresented in the literature. This limited number of studies may reflect the technical challenges associated with occlusions, complex sow–piglet interactions, and the need for high temporal precision during critical reproductive events.

In Productivity Assessment ($n = 9$), the dominant visiotype is Weight Estimation ($n = 5$), followed by Body Size Estimation ($n = 2$) and Behavior Recognition ($n = 2$). These studies primarily used 2D imaging techniques to estimate live body weight or extract morphometric features such as body length and surface area to evaluate growth

Table 4

Mapping between applications, visiotypes, included studies, computer-vision tasks and model families. A check mark (✓) indicates that the study used the corresponding CV task or model family.

Application	V/isiotype	Study	CV Task													Model family																			
			Object Detection	Classification	Segmentation	Tracking	Keypoints Detection	Feature Extraction	Pose Estimation	Regression	Counting	Action Recognition	Optical Flow	Re-identification	YOLO	Faster R-CNN	SSD	RetinaNet	DETR/RT-DETR	Mask R-CNN	U-Net	DeepLab	ResNet	VGG	EfficientNet	MobileNet	DenseNet	Inception/GoogLeNet	ViT/Transformer	HRNet	Hourglass	LSTM/RNN	Tracker (SORT-family)	Classical-MIL	Other
Activity Monitoring	Behavior Recognition	[82]	✓											✓											✓										✓
		[83]	✓			✓								✓								✓										✓			
		[84]	✓			✓								✓	✓																	✓			
		[85]					✓																✓								✓				
		[86]	✓											✓								✓													
		[87]	✓															✓				✓													
		[88]			✓	✓					✓		✓															✓							
		[89]	✓													✓	✓						✓		✓									✓	
	[90]	✓												✓																					
	Body Part Detection	[91]					✓																						✓						
	Body Size Estimation	[92]					✓															✓													
	Counting	[93]								✓																									✓
		[94]			✓									✓	✓		✓																		✓
		[95]	✓			✓								✓											✓										
		[96]									✓																								✓
		[97]			✓									✓																					
		[98]							✓														✓												
		[99]	✓												✓																				
	Counting, Tracking	[100]	✓			✓								✓																		✓			
	Detection	[47]					✓																					✓						✓	✓
		[101]	✓											✓																					
		[102]			✓		✓							✓																					
		[73]			✓																✓														✓
		[103]			✓																														✓
		[104]	✓												✓				✓																
		[105]	✓											✓																					
		[76]	✓																																✓
		[106]	✓			✓									✓																	✓		✓	
		[107]	✓												✓																				✓
		[108]	✓												✓																				
		[109]	✓												✓																				
	Detection, Posture Recognition	[110]			✓														✓	✓															✓
		[111]			✓															✓	✓														
	Detection, Tracking	[112]	✓													✓							✓									✓		✓	
		[11]	✓			✓								✓																		✓			
	Face Recognition	[113]		✓																		✓													
		[114]	✓											✓									✓		✓		✓								

(continued on next page)

Table 4 (continued)

[illegible]

(continued on next page)

(See References [11,18,40,45-170] is cited in Table body part).

performance [17,176,177]. Although weight and body size estimation are directly linked to production efficiency and economic return, the relatively limited number of studies suggests that automated productivity assessment remains less explored compared to activity and welfare monitoring.

The Health Status ($n = 8$) category incorporates a broader range of specific visiotypes. Posture Recognition ($n = 3$) is the most common, supplemented by Body Part Detection ($n = 1$), Behavior Recognition ($n = 1$), Respiration Rate Monitoring ($n = 1$), Shoulder Lesion Detection ($n = 1$), and Stress Monitoring ($n = 1$). These approaches targeted early detection of health problems, including lameness [127], respiratory irregularities [18], skin lesions [129], and physiological stress responses [128]. This may be due to the complexity of capturing subtle clinical signs through visual data alone, particularly for internal or physiological conditions that are not easily observable.

Overall, the results demonstrate that activity and behavior-related visiotypes dominate current CV applications in PLF for pigs, while more specialized physiological and health-related indicators remain comparatively underrepresented.

5.2. RQ2: What are the CV models and networks used for PLF in pig production?

Across the reviewed studies, object detection architectures constitute the dominant class of models. In particular, the YOLO family, Faster R-CNN, and SSD are the most frequently adopted frameworks. These models demonstrate robust performance under commercial barn conditions and are widely applied to tasks such as estrus detection [113], piglet tracking [171], and tail-biting detection [151]. Their popularity is largely explained by their balance between detection accuracy, inference speed, and robustness to occlusion and variable illumination.

For pixel-level analysis, instance segmentation models are predominant. Mask R-CNN and its variants are commonly used for precise body contour extraction, posture classification, and behavioral monitoring in group-housing environments, including crowded pens [120,156]. These architectures enable fine-grained spatial reasoning, which is critical for distinguishing overlapping animals and extracting posture-related visiotypes.

Individual identification is typically addressed through CNN-based image classification approaches using face or head images of pigs. Lightweight and customized CNN architectures frequently achieve accuracies exceeding 95%, enabling non-contact and scalable individual tracking [86]. More recent systems incorporate vision transformers (ViT) and multimodal transformer-based architectures, which enhance spatial-temporal modeling and improve robustness across lighting conditions and viewing angles [154].

At the task level, object detection is the most prevalent computer vision formulation in the dataset, followed by classification and

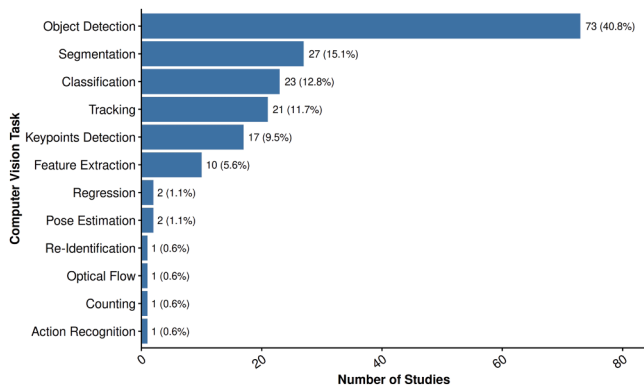


Fig. 10. Distribution of computer vision task types (e.g., detection, classification, segmentation) across the selected studies.

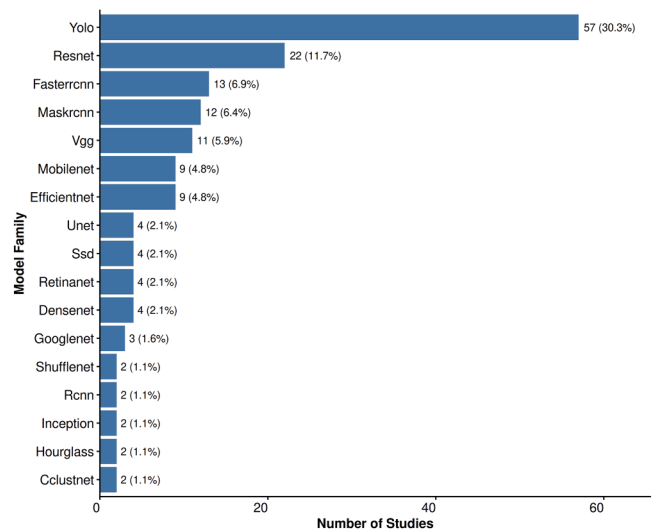


Fig. 11. Distribution of model families (e.g., YOLO, Faster R-CNN, ResNet, VGG) used in pig-focused PLF studies.

segmentation (Fig. 10). This distribution reflects the operational priorities of precision livestock farming, where identifying individuals, localizing animals, tracking movement, and detecting specific behaviors are central functional requirements.

Fig. 11 summarizes the distribution of model families. The most frequently adopted architectures are YOLO and Faster R-CNN for object detection, Mask R-CNN for instance segmentation, and ResNet and VGG for image classification. Together, these model families account for the majority of implementations across studies. In contrast, alternative architectures such as MobileNet and EfficientNet are only sporadically used and represent a minor share of the total.

Beyond the primary vision architectures used for image-based feature extraction, several studies employ additional machine learning models to analyze the extracted features. The primary models (typically CNN or related vision architectures) process individual frames to detect animals, localize body parts, or estimate postures and other visual attributes. The resulting features are then used as inputs to secondary models that perform higher-level inference, such as detecting behaviors from temporal movement patterns. As shown in Fig. 12, Long Short-Term Memory (LSTM) networks are the most frequently employed secondary models, particularly in behavior recognition pipelines where temporal dependencies are essential [144]. These recurrent architectures enable the modeling of sequential movement patterns and transitions between behavioral states.

Transformer-based architectures remain comparatively underrepresented. Although they began appearing in the literature around 2019, they account for less than 15% of all reported models (Fig. 13). This indicates an early-stage or cautious adoption of transformer paradigms in pig-specific PLF research, despite their rapid expansion and demonstrated advantages in broader computer vision domains.

5.3. RQ3: What are the evaluation metrics used for CV models for PLF in pig production?

Evaluation of computer vision (CV) models in pig systems relies on a heterogeneous set of metrics that depend on the underlying task. To improve interpretability, the reported metrics were grouped into four families: *Classification / Recognition*, *Detection / Segmentation*, *Tracking*, and *Regression / Efficiency / Pose*.

Classification and recognition studies are most commonly evaluated using Accuracy, Precision, Recall, F1-score, Specificity, Sensitivity, and AUC. These metrics are primarily used in studies addressing posture recognition, behavior recognition, identification, and related classifi-

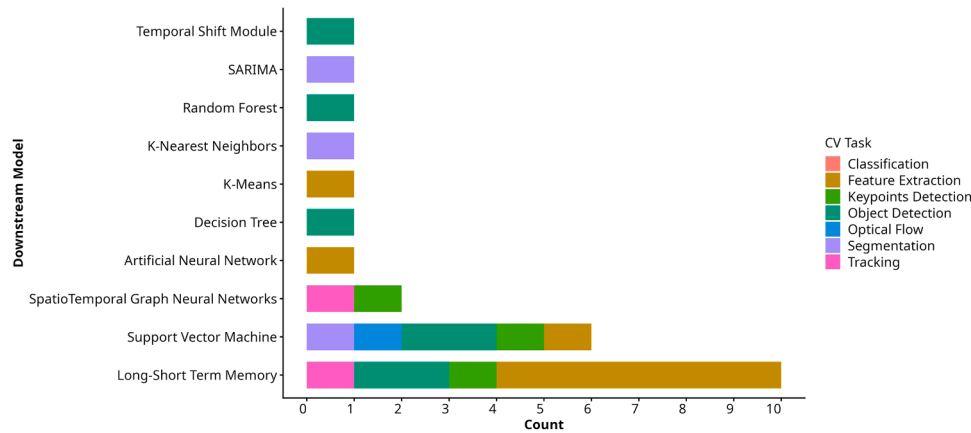


Fig. 12. Distribution of downstream models (e.g., LSTM, transformers) by CV task type.

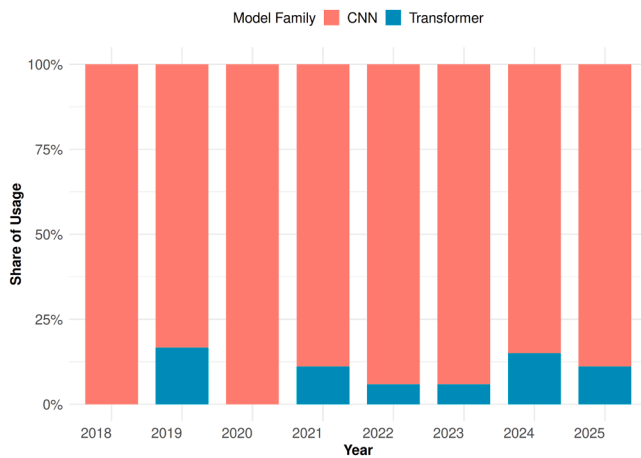


Fig. 13. Share of CNN-based versus transformer-based architectures over time.

cation tasks. Detection and segmentation studies rely mainly on overlap- and precision-based metrics such as mAP, mAP50, mAP50–95, AP50, AP75, IoU, mIoU, and MAR [92,118,144]. Multi-object tracking systems are typically assessed using MOTA, MOTP, HOTA, IDF1, and the absolute number of ID-switches [67,103]. Finally, regression and estimation tasks, such as body-weight or body-size estimation, are evaluated using error-based metrics including MAE, MSE, RMSE, and MAPE, together with correlation-based measures such as R^2 and correlation coefficients [176]. This last group also includes computational and pose-related metrics such as FPS, FLOPs, GFLOPs, PCK, and MPCK.

Two complementary aspects were examined: (1) how frequently individual metrics were reported across the included studies, and (2) the distribution of their reported numerical values.

Fig. 14 summarizes metric usage across the four groups. Within the *Classification / Recognition* family, Precision, Accuracy, and Recall are the most frequently reported metrics. In the *Detection / Segmentation* family, mAP50, mAP50–95, and mAP appear most often, indicating the predominance of modern detection benchmarks. In *Tracking*, MOTA and IDF1 are the dominant evaluation metrics, while in *Regression / Efficiency / Pose*, FPS, R^2 , RMSE, and MAE are the most frequently reported. Overall, the results suggest that evaluation practices in pig-focused PLF studies largely follow the conventions of the broader computer vision literature, but remain fragmented across task types.

Fig. 15 shows the distribution of reported values for the directly comparable performance scores (the bounded, higher-is-better metrics on a common 0–100 scale grouped by task family). These scores concentrate near the upper end of the scale, reflecting the widespread reporting of high performance. Because evaluation metrics otherwise

differ markedly in scale, direction, and units, the complete set (including error-based, runtime, computational-complexity, and count metrics that cannot share an axis) is summarized individually in Table 5, which reports for each metric its direction of improvement, the number of reporting studies, and the median and range of reported values.

5.4. RQ4: What are the challenges associated with the CV-based pig identification, detection, behavior and health monitoring?

Since barn environments are heterogeneous, robustness and generalizability are also core evaluation aspects. CV systems are therefore tested under lighting variation, occlusions, and across distinct farms or pen designs to quantify domain-shift performance and cross-farm generalization [71,178–180]. For practical on-farm deployment, additional operational metrics are considered, such as camera quality and hardware utilization [156]. Finally, dataset quality plays a critical role: studies often report inter-annotator agreement and annotation diversity.

The challenges reported across the included studies are summarized in Fig. 16. Because many studies report more than one challenge, the values are expressed as a share of all challenge occurrences ($n = 264$ mentions) rather than as a proportion of the 131 studies. Occlusion was the most frequently cited challenge (80 mentions, 30.3%), followed by lighting variability (39 mentions, 14.8%), behavioral complexity (36 mentions, 13.6%), posture-recognition difficulty (28 mentions, 10.6%), overlapping animals (23 mentions, 8.7%), pen-environment constraints (23 mentions, 8.7%), inconsistent image quality (18 mentions, 6.8%), hardware limitations (11 mentions, 4.2%), and camera-position issues (6 mentions, 2.3%). Table 6 breaks these challenges down by application (study type): occlusion predominates in activity-monitoring studies, with overlapping and pen-environment constraints likewise concentrated there (group housing), whereas behavioral complexity is most frequent in welfare studies.

These findings highlight the need for models capable of handling heterogeneous recording conditions on farms and suggest that cross-farm validation and multimodal data integration remain insufficiently explored.

5.5. A closer look at behavior recognition

Given that behavior recognition emerged as the most frequently reported visiotype (Fig. 9), we conducted a more detailed analysis of this subset of studies. In particular, we examined (i) whether the contribution was methodological or application-oriented, (ii) how this distribution evolved over time, and (iii) which behavioral domains were most frequently addressed.

To better characterize the research landscape, each study was categorized as either *method-focused* or *application-focused*. Method-focused

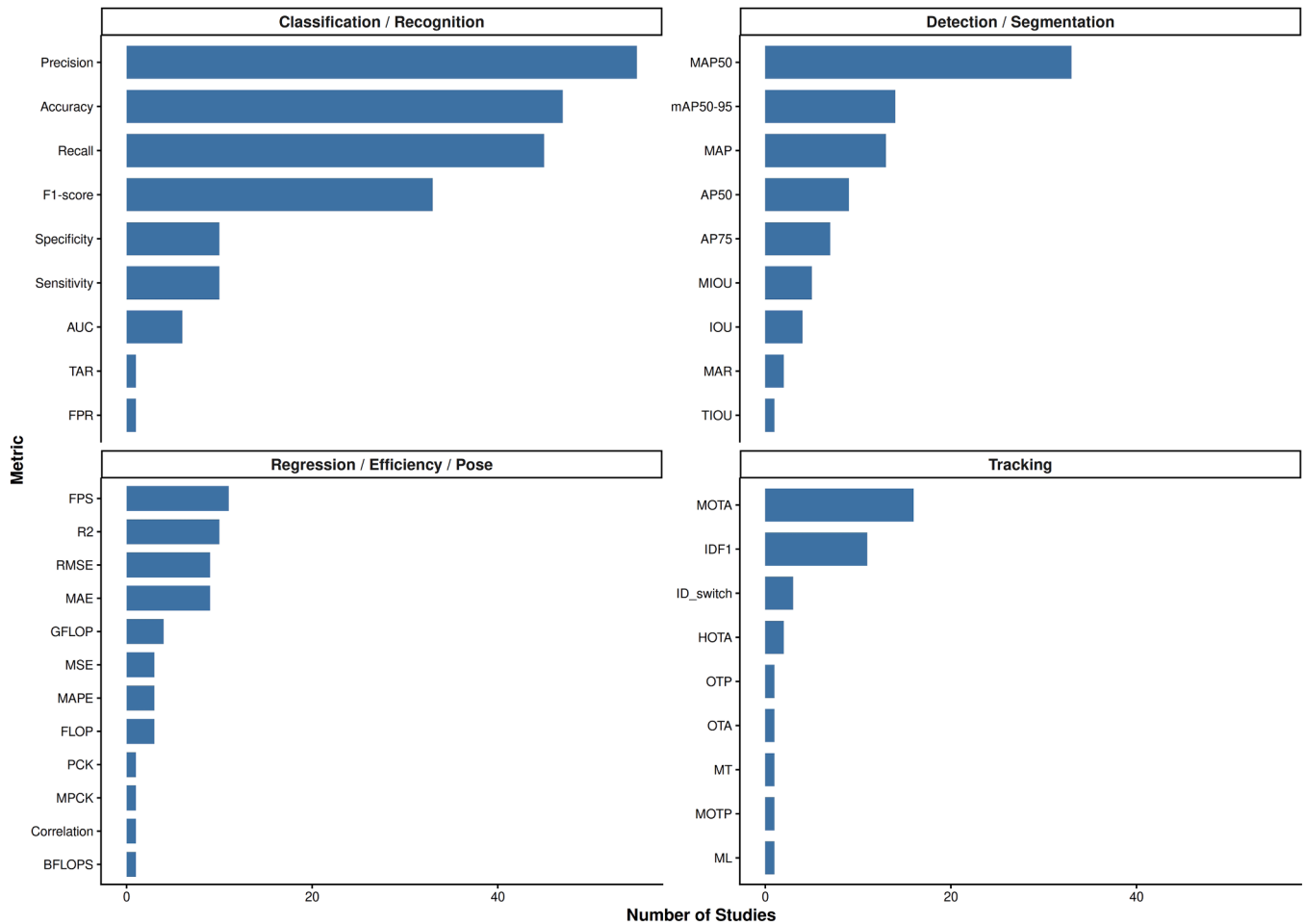


Fig. 14. Frequency of reported evaluation metrics across the included studies, grouped into four families: Classification/Recognition, Detection / Segmentation, Tracking, and Regression / Efficiency / Pose. Precision, Accuracy, and Recall are the most frequently reported classification metrics, while MAP50 and mAP-based measures dominate detection and segmentation studies.

papers propose novel architectures, substantial model adaptations, or domain-specific computer vision techniques tailored to pig production. In contrast, application-focused papers primarily adopt established architectures (e.g., YOLO, Faster R-CNN, ResNet) to automate behavioral observation tasks or to evaluate feasibility in practical farm settings. Here, application refers to the use of largely off-the-shelf models without major architectural innovation.

Fig. 17 presents the yearly distribution of both categories. From approximately 2023 onward, method-focused studies account for nearly half of all published work in behavior recognition. This shift indicates a growing investment from the engineering and computer vision communities in developing problem-specific solutions rather than relying solely on model transfer from general vision benchmarks.

The distinction between methodological and application-oriented work is further reflected in publication venues. As shown in Fig. 18, method-focused contributions are predominantly published in engineering and computer vision journals, whereas application-focused studies appear more frequently in animal science and agricultural journals. This pattern highlights a disciplinary divide between algorithmic innovation and domain-driven deployment.

Regarding behavioral targets, posture-related behaviors (e.g., lying, sitting, standing) are the most frequently studied (Fig. 19). Their dominance is likely linked to annotation simplicity and clear visual separability. Feeding-related behaviors (feeding, drinking) form the second-largest cluster. In contrast, aggressive and damaging behaviors are comparatively underrepresented, despite their high relevance for welfare assessment and herd management.

Finally, Fig. 20 maps behaviors (colored by category) to their associated application domains. Notably, postural behaviors contribute across all application types, whereas several behavior categories are confined to a single application area. For example, social and aggressive behaviors are reported exclusively in studies targeting welfare applications.

5.6. Further analysis

Beyond the primary research questions, additional aspects extracted from the review table were examined. In particular, annotation practices, an essential component of supervised computer vision (CV) model development, were analyzed. Fig. 21 presents the distribution of annotation software reported across the reviewed studies. Open-source tools were most frequently used, particularly for object detection and segmentation tasks, reflecting their accessibility and broad adoption within the research community. In contrast, comparatively fewer studies reported the use of proprietary or custom-developed annotation systems.

Reproducibility was also assessed through reported data and code availability. This aspect is particularly relevant in agricultural CV research, where datasets are frequently farm-specific and not publicly accessible. Fig. 22 summarizes the proportion of studies that shared datasets, source code, both, or neither. Only 0.7% of the reviewed papers provided both dataset and code. Approximately 5.2% shared code only, while 7.5% made datasets publicly available. The majority of studies did not provide either resource, limiting direct reproducibility, benchmarking, and independent validation.

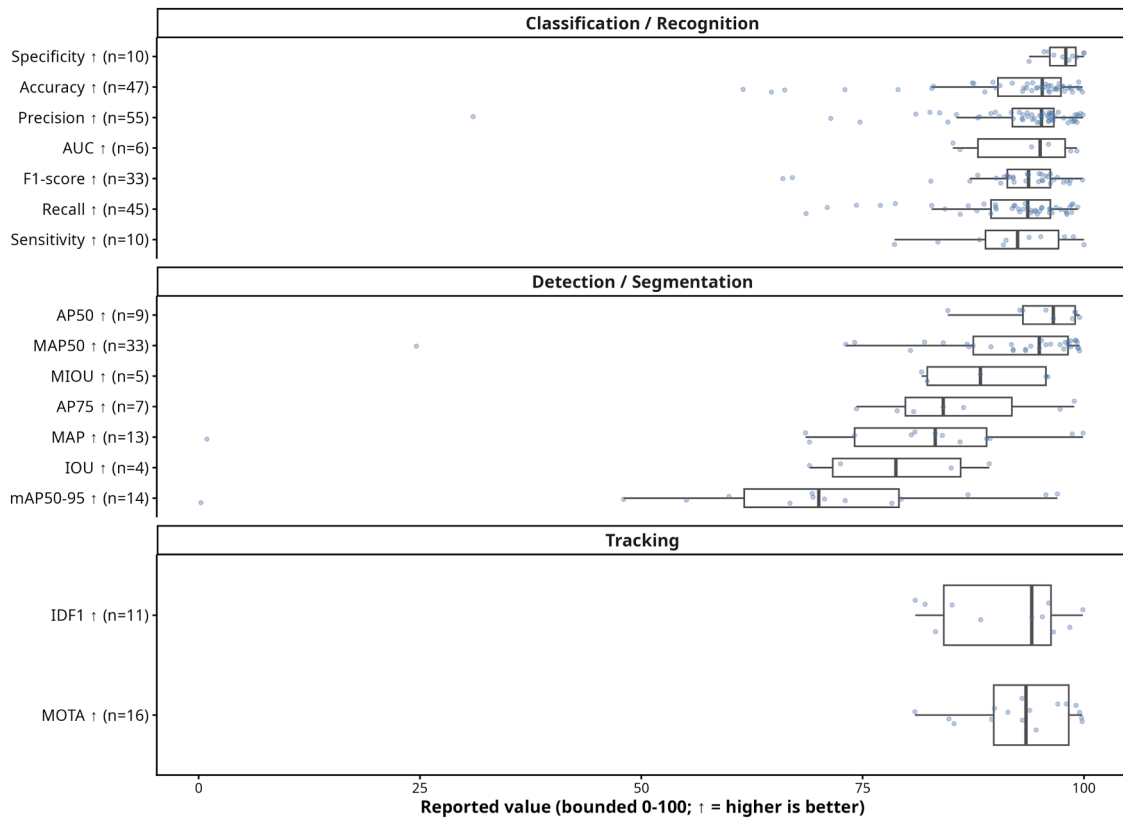


Fig. 15. Boxplots of reported evaluation metric values, grouped into four families: Classification / Recognition, Detection/Segmentation, Tracking, and Regression / Efficiency / Pose. The figure illustrates the within-metric distribution of reported values across studies; because several metrics have different scales and units, comparisons should be made within rather than across metric types.

Image resolution characteristics were additionally examined. Fig. 23 shows that 1920×1080 and 1280×720 (both widescreen formats) together with 1024×768 are the three most frequently reported resolutions across studies.

Fig. 24 presents the distribution of the number of labeled images used per study (excluding extreme video-frame outliers) and highlights the top three most common ranges. Higher image resolutions provide greater spatial detail, which is beneficial for fine-grained tasks such as segmentation; however, they also increase computational demands during training and inference.

5.7. Task-based synthesis

To complement the per-question analyses and provide a single task-oriented overview, Table 3 consolidates the reviewed studies by visio-type (task category). For each category it reports the number of studies (consistent with Fig. 9), the typical input, the annotation type, the most common model families, the most frequently reported evaluation metrics, and the main reported limitation. Annotation type is intrinsically determined by the task for primitive vision tasks (boxes for detection, masks for segmentation, keypoints for pose) and varies only for higher-level visiotypes such as behavior recognition; segmentation and pose estimation are therefore treated as cross-cutting methods (Section 5.2) rather than standalone categories. This view integrates the problem (RQ1), model (RQ2), evaluation (RQ3) and challenge (RQ4) dimensions for each task in one place.

5.8. Validation practices

To complement the reported metrics, we assessed the validation design of every included study (Table 7). A train/validation/test split was reported by 125 studies (95%), but a genuinely independent test set

(one not composed of random frames from the same videos or animals used for training) could be confirmed in only 58 (44%); for a further 48 the split was too under-described to verify. Cross-validation was used in 13 studies and temporal validation in only 15; the remaining 103 used neither. Most importantly, external or cross-farm validation was reported in just 17 studies (13%): 93 drew all data from a single source and 91 from a single site. We accordingly judged the risk of data leakage to be high in 45 studies, most often from random frame-level splitting of continuous video or from augmentation applied before splitting. This weakness is common to every task category (Table 7): external/cross-farm validation never exceeded 28% (tracking) and was absent for classification, while data-leakage risk was high in roughly a third or more of detection, classification, and segmentation studies. These patterns indicate that the predominantly high metrics summarized in Table 5 are typically obtained under within-dataset, single-site conditions and may overstate real-world, cross-farm performance.

6. Discussion of gaps and promising future research directions

The synthesis of the reviewed literature highlights substantial methodological progress in DL-based computer vision for pig production, but also reveals structural limitations that constrain cumulative scientific advancement and large-scale deployment.

6.1. Data availability, representativeness, and standardization

A primary bottleneck concerns data availability, representativeness, and standardization. As summarized previously, most studies rely on datasets collected in single facilities under relatively homogeneous housing and environmental conditions. Such restricted sampling limits external validity and cross-farm generalization. Moreover, annotation practices vary considerably, both in labeling granularity and behavioral

Table 5

Summary of reported evaluation metrics across the included studies, grouped by metric family. Metrics are summarized individually because they differ in scale, direction and interpretation and cannot be compared across families. $\uparrow$ = higher is better, $\downarrow$ = lower is better. n = number of studies reporting the metric; Median and Range are over the reported values (Range shown as “—” when $n = 1$). Bounded performance scores lie on a 0–100 scale; a few studies reported mAP, mAP50–95, FPR or correlation on a 0–1 scale, left as reported (hence the low minima). ID-switches are absolute counts; FLOPs/GFLOPs/BFLOPs are computational-complexity measures and FPS is inference speed, so none are comparable to the 0–100 scores. MAPE includes one high outlier (82.4 %); the single PCK value (11.4) is unusually low and is reported as found.

Metric	Direction	n	Median	Range (min–max)
Classification / Recognition				
Precision	$\uparrow$	55	95.2	31.0–99.9
Accuracy	$\uparrow$	47	95.3	61.5–99.8
Recall	$\uparrow$	45	93.7	68.6–99.3
F1-score	$\uparrow$	33	93.7	66.0–99.8
Sensitivity	$\uparrow$	10	92.5	78.6–100.0
Specificity	$\uparrow$	10	98.0	93.8–100.0
AUC	$\uparrow$	6	95.0	85.2–99.2
TAR	$\uparrow$	1	99.7	—
FPR	$\downarrow$	1	0.10	—
Detection / Segmentation				
mAP50	$\uparrow$	33	95.0	24.6–99.5
mAP50–95	$\uparrow$	14	70.0	0.26–97.0
mAP	$\uparrow$	13	83.2	0.93–99.9
AP50	$\uparrow$	9	96.5	84.6–99.5
AP75	$\uparrow$	7	84.1	74.3–98.9
mIoU	$\uparrow$	5	88.3	81.7–95.9
IoU	$\uparrow$	4	78.8	69.0–89.3
MAR	$\uparrow$	2	79.9	73.4–86.4
TIoU	$\uparrow$	1	96.8	—
Tracking				
MOTA	$\uparrow$	16	93.5	80.9–99.8
IDF1	$\uparrow$	11	94.1	80.9–99.9
HOTA	$\uparrow$	2	76.3	67.9–84.7
MOTP	$\uparrow$	1	84.1	—
OTA	$\uparrow$	1	92.7	—
OTP	$\uparrow$	1	68.8	—
ID-switches	$\downarrow$	3	24	18–165 (count)
Pose				
MPCK	$\uparrow$	1	95.3	—
PCK	$\uparrow$	1	11.4	— (low; verify)
Error / regression (lower is better)				
MAE	$\downarrow$	9	1.25	0.43–6.01
RMSE	$\downarrow$	9	2.13	0.85–7.67
MSE	$\downarrow$	3	0.32	0.08–4.45
MAPE	$\downarrow$	3	2.82	1.97–82.41 (%)
Agreement / fit (higher is better)				
R ²	$\uparrow$	10	86.3	67.0–99.3
Correlation	$\uparrow$	1	0.92	—
Inference speed				
FPS	$\uparrow$	11	29.0	6.4–75.0
Computational complexity (lower is better)				
GFLOPs	$\downarrow$	4	9.5	4.9–25.2
FLOPs	$\downarrow$	3	78.7	8.0–88.8
BFLOPs	$\downarrow$	1	4.96	—

definitions (see Table 2). The distribution of annotation tools (Fig. 21) further indicates reliance on heterogeneous workflows, often without reported inter-annotator agreement. This fragmentation complicates reproducibility and benchmarking. The limited availability of shared datasets and code (Fig. 22) reinforces this concern, despite increasing recognition of standardized PLF datasets as a prerequisite for progress [16]. Future research should prioritize the creation of openly accessible, collaborative, multi-site benchmark datasets with harmonized behavioral ontologies and clearly documented annotation protocols as described in [181].

6.2. Architectures and representation learning

From a methodological perspective, the dominance of CNN is evident

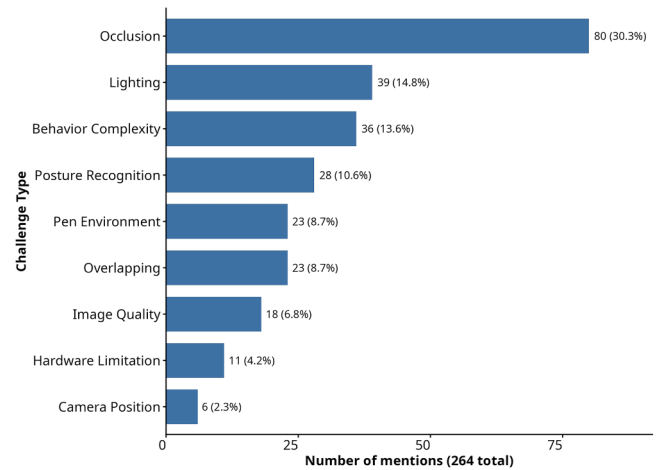


Fig. 16. Distribution of technical and environmental challenges reported in the reviewed studies. Percentages are computed over the total number of challenge occurrences across all studies ($n = 264$), not over the number of included studies; because a study may report multiple challenges, the counts sum to more than 131 and the percentages sum to 100%.

(Fig. 13), with YOLO-based detectors and CNN classifiers constituting the prevailing backbone architectures (Fig. 11). While these approaches have demonstrated robust performance for detection, counting, and posture classification tasks, more recent representation learning paradigms remain comparatively underexplored. Transformer-based models and hybrid CNN–Transformer architectures (Fig. 13), which have shown strong performance in general vision tasks, are only beginning to emerge in pig-related applications [33,154]. Given their capacity for global context modeling and long-range dependency capture, they are particularly promising for complex group-level behavior analysis. In addition, self-supervised and foundation-model approaches, increasingly discussed in broader AI-for-livestock contexts [178,182], may reduce reliance on fully annotated datasets and mitigate one of the central cost drivers in PLF vision research.

6.3. Task scope and behavioral modeling

Task scope also reveals an imbalance. As illustrated in Fig. 10, foundational perception tasks (e.g., detection, segmentation, tracking) substantially outnumber higher-level behavioral inference tasks. Although aggression detection and tail-biting prediction are represented [40,153], longitudinal modeling of social dynamics, dominance hierarchies, and early-stage welfare deterioration remains limited. Many studies operate at frame-level classification, whereas socially meaningful events are intrinsically temporal and relational. Future work should therefore emphasize spatiotemporal modeling frameworks, including sequence-based architectures and interaction-aware models capable of capturing inter-animal relationships over extended time horizons.

6.4. Evaluation and validation practices

Evaluation practices further limit comparability. As shown in Figs. 14 and 15, performance metrics are reported inconsistently across studies, with varying use of accuracy, F1-score, mAP, and task-specific indicators. Cross-study comparison is therefore difficult, and external validation across farms is rare. Indeed, only 17 of the 131 studies (13%) reported external or cross-farm validation (Table 7), and we identified a high risk of data leakage in 45. Short evaluation windows and limited environmental variability reduce confidence in real-world robustness. Standardized benchmarking protocols and cross-site validation designs are essential to ensure that reported performance translates to commercial deployment. These gaps are, moreover, unevenly distributed

Table 6
Reported challenges by application (type of study). Values are the number of included studies reporting each challenge within each application. Total is the number of distinct studies reporting the challenge (matching Fig. 16). Because a study may serve more than one application, a row's cells can sum to more than its Total (same convention as Fig. 9).

Challenge	Activity Monitoring	Welfare	Breeding & Reproduction	Productivity Assessment	Health Status	Total
Occlusion	41	23	9	5	5	80
Lighting	23	11	3	2	1	39
Behavior complexity	7	22	4	3	2	36
Posture recognition	12	9	2	4	2	28
Overlapping	15	5	2	0	1	23
Pen environment	16	3	4	0	0	23
Image quality	8	3	3	4	1	18
Hardware limitation	7	3	1	1	0	11
Camera position	1	2	1	1	2	6

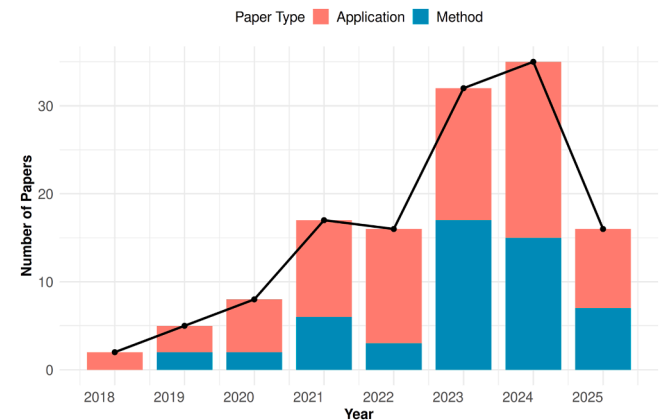


Fig. 17. Yearly distribution of method-focused versus application-focused studies.

across tasks in ways that reflect underlying data constraints rather than chance. Tracking shows the highest rate of external or cross-farm validation (28%; Table 7), because multi-object tracking is conventionally benchmarked on held-out video sequences and its core value, preserving identity across scenes, can only be demonstrated on footage from new

pens or farms. Classification studies, by contrast, reported no cross-farm validation at all (0%), because posture- and behavior-recognition datasets are typically built by sampling additional frames from the same fixed cameras, which lowers annotation cost but discourages sourcing an independent site. Pose estimation carries the highest data-leakage risk (60%): keypoint annotation is the most labor-intensive labeling task, so datasets are small and drawn from few animals, and the resulting reliance on dense frame sampling from continuous video, combined with random frame-level splitting, places near-duplicate frames in both the training and test sets. More broadly, annotation-cheap and data-rich tasks such as detection and classification gravitate toward large within-dataset splits, whereas annotation-expensive or less mature tasks such as pose estimation and counting are more exposed to leakage, so validation rigor tracks annotation cost, data availability, and task maturity rather than being randomly distributed.

6.5. Multimodal and vision-language integration

While the present review is restricted to 2D computer vision approaches based on standard RGB imagery from visible-light cameras, it is noteworthy that none of the surveyed studies adopt multimodal foundation models in the contemporary sense of vision-language integration. All identified approaches operate within purely visual pipelines, mapping images directly to task-specific outputs without leveraging

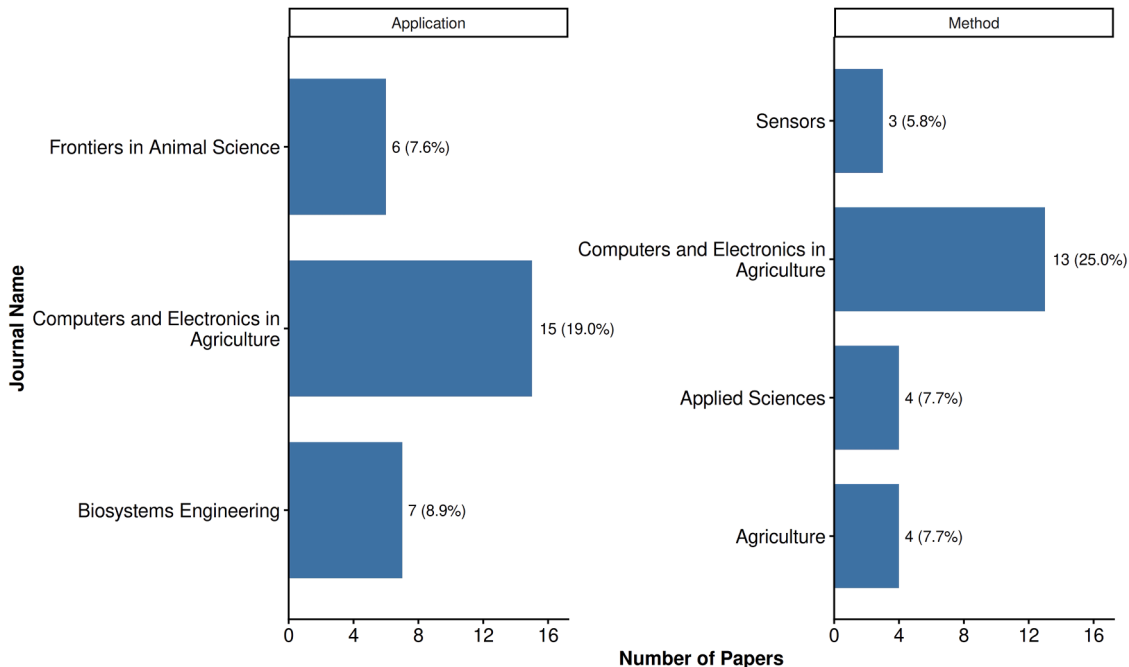


Fig. 18. Comparison of method- versus application-focused papers across the journals appearing more than three times in the reviewed studies.

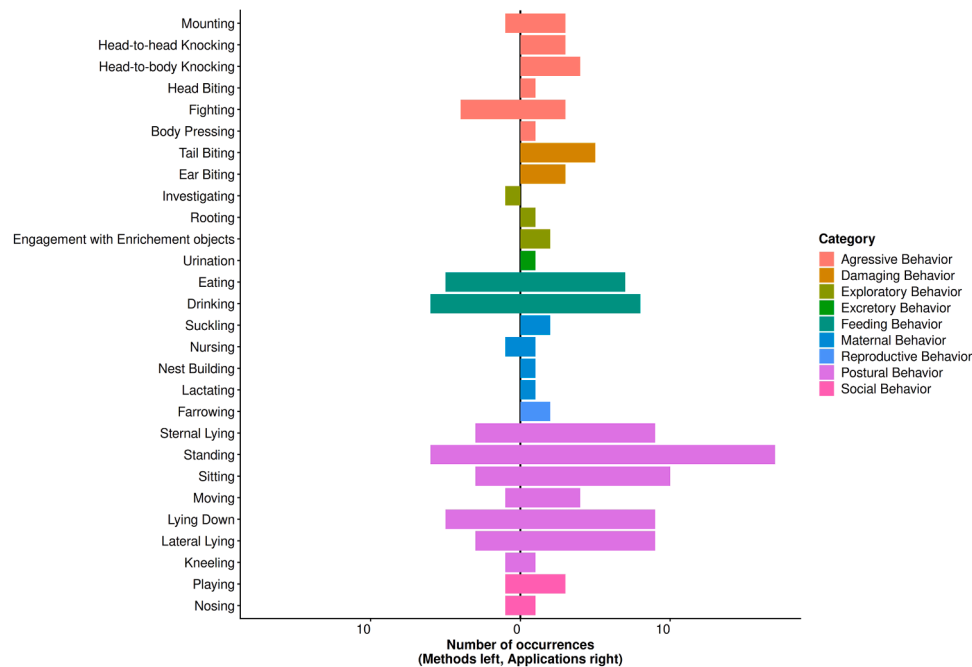


Fig. 19. Distribution of studied behaviors across method- and application-focused papers.

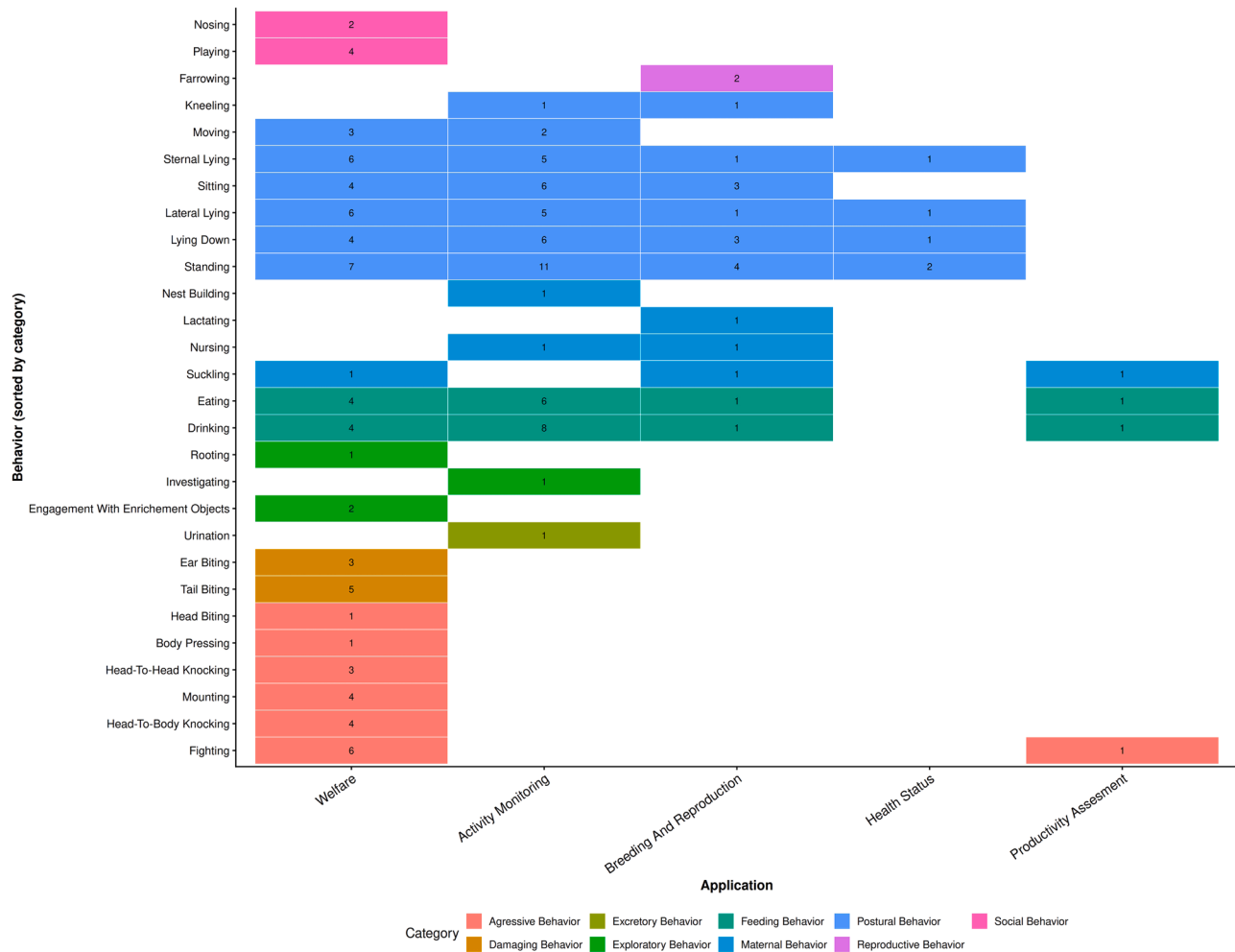


Fig. 20. Distribution of behaviors and their associated applications across the reviewed studies.

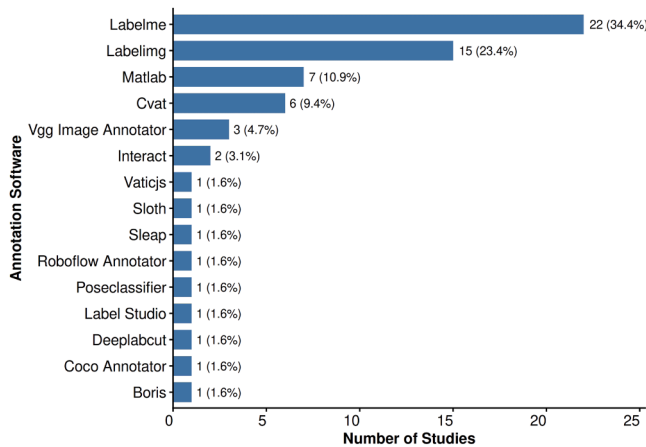


Fig. 21. Distribution of annotation software used in the reviewed studies.

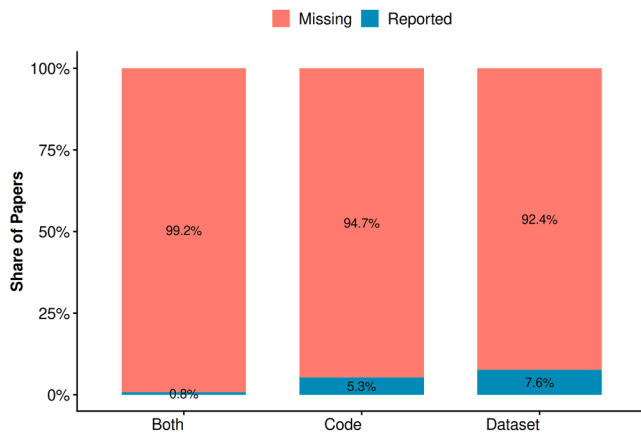


Fig. 22. Share of studies offering public datasets, code, both, or neither.

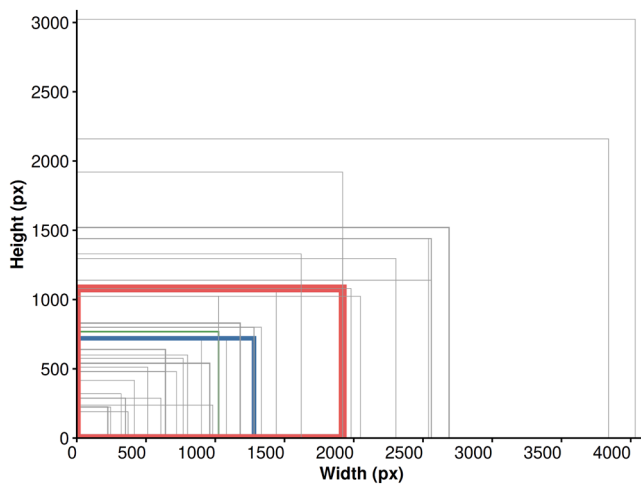


Fig. 23. Distribution of the most frequently reported image resolutions.

textual context, semantic priors, or cross-modal alignment mechanisms. Recent advances in large vision-language models (LVLMs) have demonstrated strong generalization capabilities through joint visual-text representation learning and instruction tuning. The absence of such architectures in RGB-based pig monitoring represents a conceptual gap rather than a sensor-level limitation. Future research could explore the integration of vision-language models to enable semantically richer

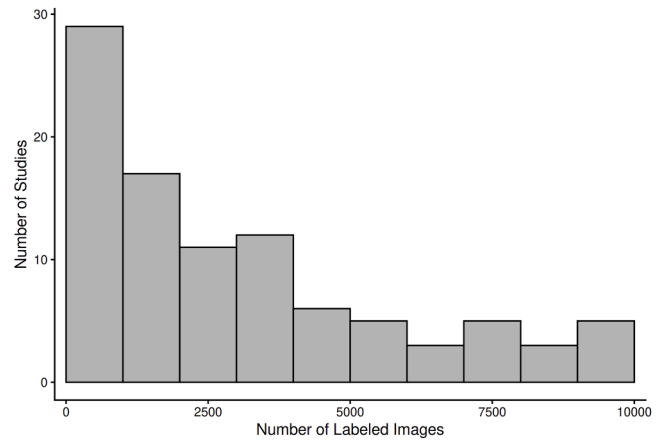


Fig. 24. Distribution of the number of labeled images used per study (excluding video-frame outliers).

behavioral interpretation, zero-shot or few-shot task adaptation, and more flexible query-based monitoring systems (e.g., prompting models with welfare-relevant descriptions). Investigating whether foundation models pretrained on large-scale natural image-text corpora can be effectively adapted to livestock production environments constitutes a promising direction.

6.6. On-farm deployment and practical value

Finally, aspects related to practical on-farm deployment remain insufficiently addressed and are rarely connected to reported performance. The challenges summarized in Fig. 16 and Table 6 are precisely those that erode field performance: occlusion and overlapping animals in group housing (the most frequently cited challenges) reduce detection and tracking reliability and drive identity switches, while inconsistent image quality, camera-position constraints, and variable lighting further limit robustness. Several practically critical factors, however, are reported only sporadically: motion blur, lens soiling, and changes in pig appearance across growth stages (all of which affect long-term accuracy) are seldom evaluated. Indicators of real-time and edge feasibility are likewise scarce: inference speed is reported by only a minority of studies (median ≈ 29 FPS; Table 5), and computational-complexity measures (FLOPs/GFLOPs), on-device or embedded-hardware testing, maintenance requirements, and economic or cost-benefit analyses are reported even less often. Consequently, the suitability of most reviewed models for continuous, low-latency operation on commercial farms cannot yet be judged. Moving beyond algorithm-centric evaluation toward holistic system validation (linking robustness under occlusion, lighting, and crowding to real-time performance, edge-deployment cost, and long-term operational stability) will be essential for translating these methods into practice.

6.7. Summary and research priorities

In summary, advancing DL-based computer vision in pig production is best understood not as a flat set of parallel goals but as three inter-dependent layers. At the base lie the enabling conditions: standardized, openly accessible multi-site datasets and harmonized evaluation and benchmarking protocols. These are prerequisites rather than parallel items, because without shared benchmarks and leakage-aware validation the field cannot reliably tell whether a methodological advance represents genuine progress. Built on this base, the methodological layer, namely broader adoption of transformer and self-supervised or foundation-model architectures together with deeper modeling of temporal and social behavioral dynamics and systematic multimodal integration, can be developed and, crucially, compared on a common

Table 7

Validation strategy of the included studies, by primary computer-vision task. Cells give the number of studies and, in parentheses, the percentage of the task's n. "Independent test set" = a test set not composed of random frames from the same videos/animals used for training; "External / cross-farm" = evaluated on a different farm, site, or dataset than used for training; "Single-site" = all data from a single site/farm; "High leakage risk" = our assessment, most often arising from random frame-level splitting of continuous video or from augmentation applied before splitting. Values come from full-text reading of all 131 studies and are coded conservatively: not-reported or under-described cases are not counted as positive. "Task" is each study's primary CV task; the final row is all 131 studies.

Task	n	Split reported	Independent test set	Cross-validation	Temporal validation	External / cross-farm	Single-site	High leakage risk
Detection	44	42 (95%)	20 (45%)	2 (5%)	5 (11%)	7 (16%)	32 (73%)	15 (34%)
Classification	29	28 (97%)	17 (59%)	6 (21%)	5 (17%)	0 (0%)	22 (76%)	9 (31%)
Segmentation	18	18 (100%)	3 (17%)	0 (0%)	0 (0%)	2 (11%)	13 (72%)	7 (39%)
Tracking	18	15 (83%)	12 (67%)	3 (17%)	5 (28%)	5 (28%)	11 (61%)	3 (17%)
Pose	10	10 (100%)	3 (30%)	0 (0%)	0 (0%)	2 (20%)	7 (70%)	6 (60%)
Counting	8	8 (100%)	2 (25%)	1 (12%)	0 (0%)	1 (12%)	3 (38%)	4 (50%)
Other	4	4 (100%)	1 (25%)	1 (25%)	0 (0%)	0 (0%)	3 (75%)	1 (25%)
All studies	131	125 (95%)	58 (44%)	13 (10%)	15 (11%)	17 (13%)	91 (69%)	45 (34%)

footing. The translation layer then converts validated methods into field-ready systems: deployment under commercial conditions, real-time and edge feasibility, maintenance, and economic or cost-benefit assessment. Progress at each layer is gated by the one beneath it, since architectural innovation yields little durable value until it can be benchmarked, and benchmarked methods deliver impact only once they are translated to the farm. Framing open benchmarks and evaluation protocols as the enabling base, methodological advances as the middle layer, and deployment and economics as the downstream translation layer clarifies where coordinated investment is most needed to move from technically promising prototypes to robust, field-ready precision livestock farming solutions.

7. Synthesis: computer vision as an operationalization of visiotypes

Taken together, the reviewed studies demonstrate that computer vision in pig precision livestock farming functions primarily as a technological framework for extracting and operationalizing visiotypes as defined in Section 4. Across applications, CV systems do not measure welfare or productivity directly; rather, they quantify algorithmic outputs, such as Detection, Tracking, Behavior Recognition, Posture Recognition, Body Size Estimation, Estrus Detection, or Weight Estimation, which serve as proxies for biological state, performance, and welfare conditions.

The distribution of visiotypes across applications (Fig. 9) reveals a clear structural pattern. Foundational perception-level visiotypes (e.g., Detection and Tracking) dominate Activity Monitoring, forming the computational basis for higher-level inference. Behavioral and posture-based visiotypes are central to Welfare and Health Status applications, reflecting the reliance on observable movement and interaction patterns as indicators of animal condition. In contrast, morphometric visiotypes such as Weight Estimation and Body Size Estimation are primarily associated with Productivity Assessment.

This layered structure suggests that current CV systems in pig production follow a hierarchical abstraction pipeline: low-level spatial localization (Detection), temporal association of individuals (Tracking), recognition of body postures (Posture Recognition), and finally higher-level behavioral interpretation (Behavior Recognition). The biological relevance of these systems therefore emerges not from the raw visual input itself, but from the translation of algorithmic outputs into measurable visiotypic indicators that can support monitoring and management decisions. Table 3 operationalizes this view, summarizing for each visiotype the typical input, annotation, models, metrics and limitations.

Supplementary materials: Information extracted from the reviewed papers is provided in the Supplementary Materials as an Excel file. The code used to reproduce the figures is publicly available on GitHub: <https://github.com/nasserdr/SystematicLiteratureReview-CV-Pigs-2025>

Statement of human and animal rights*

Title: Deep Learning based 2D Computer Vision in Precision Livestock Farming for Pigs: A Systematic Literature Review and Future Research Directions

The author(s) state(s) that this research was conducted in accordance with the Helsinki Declaration as revised in 2008.

In all studies involving animals, the author(s) followed the guidelines for the use and care of laboratory animals of the author's institution or the National Research Council or any national law pertaining to care of animals in research.

Name of researcher: Hassan-Roland Nasser.

Signed: Hassan-Roland Nasser.

Date: 01.05.2026.

Name of researcher: Claudia Kasper.

Signed: Claudia Kasper.

Date: 01.05.2026.

Name of researcher: Vladimir Živković.

Signed: Vladimir Živković.

Date: 01.05.2026.

* Ethical considerations in the conduct and reporting of research: Protection of human subjects and animals in research

When reporting experiments on human subjects, authors should indicate whether the procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008. If doubt exists whether the research was conducted in accordance with the Helsinki Declaration, the authors must explain the rationale for their approach and demonstrate that the institutional review body explicitly approved the doubtful aspects of the study. When reporting experiments on animals, authors should indicate whether the institutional and national guidelines for the care and use of laboratory animals were followed.

CRediT authorship contribution statement

Hassan-Roland Nasser: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Claudia Kasper:** Writing – review & editing, Visualization, Validation, Supervision. **Vladimir Živković:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis, Data curation.

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.

Acknowledgments

This project was partially supported by the Swiss National Science Foundation (SNSF) under grant number IZSEZO_232654. We also extend our sincere thanks to Ana Gomes for creating and providing video data used to produce some of the images included in this manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.atech.2026.102426](https://doi.org/10.1016/j.atech.2026.102426).

Data availability

No data was used for the research described in the article.

References

- [1] Y. Bengio, Y. Lecun, G. Hinton, Deep learning for AI, *Commun. ACM* 64 (2021) 58–65.
- [2] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (2015) 436–444.
- [3] Y. Lu, S. Young, A survey of public datasets for computer vision tasks in precision agriculture, *Comput. Electron. Agric.* 178 (2020) 105760.
- [4] A. Dhillon, G K Verma, Convolutional neural network: a review of models, methodologies and applications to object detection, *Prog. Artif. Intell.* 9 (2020) 85–112.
- [5] D A Borges Oliveira, L G Ribeiro Pereira, T. Bresolin, R E Pontes Ferreira, J R Reboucas Dorea, A review of deep learning algorithms for computer vision systems in livestock, *Livest. Sci.* 253 (2021) 104700.
- [6] A. Nasirahmadi, S A Edwards, S M Matheson, B. Sturm, Using automated image analysis in pig behavioural research: assessment of the influence of enrichment substrate provision on lying behaviour, *Appl. Anim. Behav. Sci.* 196 (2017) 30–35.
- [7] C. Okinda, I. Nyalala, T. Korohou, C. Okinda, J. Wang, T. Achieng, P. Wamalwa, T. Mang, M. Shen, A review on computer vision systems in monitoring of poultry: a welfare perspective, *Artif. Intell. Agric.* 4 (2020) 184–208.
- [8] J. Rushen, N. Chapinal, A de Passillé, Automated monitoring of behavioural-based animal welfare indicators, *Anim. Welf.* 21 (2012) 339–350.
- [9] S. Marchegiani, G. Gislon, R. Marino, M. Caroprese, M. Albenzio, W E Pinchak, G E Carstens, L. Ledda, M F Trombetta, A. Sandrucci, M. Pasquini, P A Deligios, S. Ceccobelli, Smart technologies for sustainable pasture-based ruminant systems: a review, *Smart Agric. Technol.* 10 (2025) 100789.
- [10] M. Kashiha, C. Bahr, S. Ott, C P H Moons, T A Niewold, F O Ödberg, D. Berckmans, Automatic identification of marked pigs in a pen using image pattern recognition, *Comput. Electron. Agric.* 93 (2013) 111–120.
- [11] L E van der Zande, O. Guzhva, T B Rodenburg, Individual detection and tracking of group-housed pigs in their home pen using computer vision, *Front. Anim. Sci.* (2021).
- [12] K. Wurtz, I. Camerlink, R B D'Eath, A P Fernández, T. Norton, J. Steibel, J. Siegford, Recording behaviour of indoor-housed farm animals automatically using machine vision technology: a systematic review, *PLoS. One* 14 (2019) e0226669.
- [13] Shumaly M, Park Y, Agha S, Pandey S and Steibel J, 2025. Computer Vision Based Phenotyping and Analysis in Presence of Uncertain Identification.
- [14] A. Rohan, M S Rafiq, J. Hasan Md, F. Asghar, A K Bashir, T. Dottorini, Application of deep learning for livestock behaviour recognition: a systematic literature review, *Comput. Electron. Agric.* 224 (2024) 109115.
- [15] Scott S D, Abbas Z J, Ellid F, Dykhne E-H, Islam M M, Ayad W, Kacmorova K, Tulpan D and Gong M, 2024. Systematic literature review of vision-based approaches to outdoor livestock monitoring with lessons from wildlife studies.
- [16] A. Bhujel, Y. Wang, Y. Lu, D. Morris, M. Dangel, A systematic survey of public computer vision datasets for precision livestock farming, *Comput. Electron. Agric.* 229 (2025) 109718.
- [17] R. Dohmen, C. Catal, Q. Liu, Computer vision-based weight estimation of livestock: a systematic literature review, *N. Z. J. Agric. Res.* 65 (2022) 227–247.
- [18] Wang M, Li X, Larsen M L V, Liu D, Rault J-L and Norton T, 2023. A computer vision-based approach for respiration rate monitoring of group-housed pigs computers and electronics in agriculture.
- [19] A. Arora, A. Kumar, HOG and SIFT transformation algorithms for the underwater image fusion, in: 2021 IEEE International Conference on Technology, Research, and Innovation for Betterment of Society (TRIBES), 2021, pp. 1–5, 2021 IEEE International Conference on Technology, Research, and Innovation for Betterment of Society (TRIBES).
- [20] F. Shen, H. Liu, C. Xu, L. Ouyang, J. Zhang, Y. Chen, Y. He, VGGNet and attention mechanism-based image quality assessment algorithm in symmetry edge intelligence systems, *Symmetry* 17 (2025) 331.
- [21] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- [22] M. Tan, Q. Le, EfficientNet: rethinking model scaling for convolutional neural networks, in: Proceedings of the 36th International Conference on Machine Learning, International Conference on Machine Learning (PMLR), 2019, pp. 6105–6114.
- [23] G. Huang, Z. Liu, L. Van Der Maaten, K Q Weinberger, Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- [24] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A C Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, *Int. J. Comput. Vis.* 115 (2015) 211–252.
- [25] Y. Wang, Y. Deng, Y. Zheng, P. Chattopadhyay, L. Wang, Vision transformers for image classification: a comparative survey, *Technologies* 13 (2025) 32.
- [26] Simonyan K and Zisserman A, 2015. Very deep convolutional networks for large-scale image recognition.
- [27] C. Rao, Y. Liu, Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization, *Comput. Mater. Sci.* 184 (2020) 109850.
- [28] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, *Adv. Neural Inf. Process. Syst.* 28 (2015).
- [29] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, IEEE, 2016, pp. 779–788, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- [30] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C-Y Fu, A C Berg, SSD: single shot multibox detector, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), *Computer Vision – ECCV 2016*, Cham, Springer International Publishing, 2016, pp. 21–37.
- [31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, AN Gomez, Kaiser Ł, I. Polosukhin, Attention is all you need, in: *Proceedings of the 31st International Conference on Neural Information Processing Systems NIPS'17*, Red Hook, NY, USA, Curran Associates Inc, 2017, pp. 6000–6010.
- [32] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-end object detection with transformers, in: *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I*, Berlin, Heidelberg, Springer-Verlag, 2020, pp. 213–229.
- [33] I. Robinson, P. Robicheaux, M. Popov, D. Ramanan, N. Peri, RF-DETR: neural architecture search for real-time detection transformers, in: *The Fourteenth International Conference on Learning Representations*, 2025.
- [34] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W M Wells, A F Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham, Springer International Publishing, 2015, pp. 234–241.
- [35] H. Zeng, S. Peng, D. Li, Deeplabv3+ semantic segmentation model based on feature cross attention mechanism, *J. Phys.: Conf. Ser.* 1678 (2020) 012106.
- [36] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (2020) 386–397.
- [37] Z. Cao, G. Hidalgo, T. Simon, S. Wei, Y. Sheikh, OpenPose: realtime multi-person 2D pose estimation using part affinity fields, *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (2021) 172–186.
- [38] Wang J, Sun K, Cheng T, Jiang B, Deng C, Zhao Y, Liu D, Mu Y, Tan M, Wang X, Liu W and Xiao B, 2020. Deep high-resolution representation learning for visual recognition.
- [39] A. Mathis, P. Mamidanna, K M Cury, T. Abe, V N Murthy, M W Mathis, M. Bethge, DeepLabCut: markerless pose estimation of user-defined body parts with deep learning, *Nat. Neurosci.* 21 (2018) 1281–1289.
- [40] C. Chen, W. Zhu, J. Steibel, J. Siegford, K. Wurtz, J. Han, T. Norton, Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory, *Comput. Electron. Agric.* 169 (2020) 105166.
- [41] X. Shi, Z. Chen, H. Wang, D-Y Yeung, W. Wong, W. Woo, Convolutional LSTM Network: a machine learning approach for precipitation nowcasting, in: *Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 1 NIPS'15*, Cambridge, MA, USA 1, MIT Press, 2015, pp. 802–810.
- [42] M J Page, J E McKenzie, P M Bossuyt, I. Boutron, T C Hoffmann, C D Mulrow, L. Shamseer, J M Tetzlaff, E A Akl, S E Brennan, R. Chou, J. Glanville, J M Grimshaw, A. Hróbjartsson, M M Lahu, T. Li, E W Loder, E. Mayo-Wilson, S. McDonald, L A McGuinness, L A Stewart, J. Thomas, A C Tricco, V A Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, *BMJ* 372 (2021) n71.
- [43] B A Kitchenham, Systematic review in software engineering: where we are and where we should be going, in: *Proceedings of the 2nd International Workshop on Evidential Assessment of Software Technologies EAST '12*, New York, NY, USA, Association for Computing Machinery, 2012, pp. 1–2.
- [44] Critical Appraisal Skills Programme, 2024. CASP: Critical Appraisal Skills Programme.
- [45] Z. Liu, J. Liu, Y. Zhou, Z. Wang, A method for pig behavior target detection based on the lightweight RTMDet model, in: 2024 International Symposium on Digital Home (ISDH), 2024, pp. 25–30, 2024 International Symposium on Digital Home (ISDH).
- [46] L. Bergamini, S. Pini, A. Simoni, R. Vezzani, S. Calderara, R B D Eath, R B Fisher, Extracting accurate long-term behavior changes from a large pig dataset, in:

- VISIGRAPP 2021 - Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2021.
- [47] A. Bhujel, E. Arulmozhi, B.-E. Moon, H.-T. Kim, Deep-learning-based automatic monitoring of pigs' physico-temporal activities at different greenhouse gas concentrations, *Animals* (2021) 3089.
 - [48] C. Chen, W. Zhu, J. Steibel, J. Siegford, J. Han, T. Norton, Classification of drinking and drinker-playing in pigs by a video-based deep learning method, *Biosyst. Eng.* 196 (2020) 1–14.
 - [49] Anon enhanced swine behavior detection with YOLOs and a mixed efficient layer aggregation network in real time.
 - [50] L. Shi, H. Ying, P. Yang, C. Yang, Pig behavior detection model based on improved RT-DETR, in: 2024 6th International Conference on Electronic Engineering and Informatics, EEI 2024, 2024.
 - [51] B. Tangirala, I. Bhandari, D. Laszlo, D. K. Gupta, R. M. Thomas, D. Arya, Livestock monitoring with transformer (STARFORMER), in: 32nd British Machine Vision Conference, BMVC 2021, 2021.
 - [52] Y. Zhang, J. Cai, D. Xiao, Z. Li, B. Xiong, Real-time sow behavior detection based on deep learning, *Comput. Electron. Agric.* (2019) 104884.
 - [53] K. Zhang, Y. Zhang, H. Xu, Lightweight domestic pig behavior detection based on YOLOv8, *Appl. Sci.* 15 (2025) 6340.
 - [54] E. Psota, M. Mittek, L. Perez, T. Schmidt, B. Mote, Multi-pig part detection and association with a fully-convolutional network, *SENSORS* (2019) 852.
 - [55] W. Gorssen, C. Winters, R. Meyermans, R. D'Hooge, S. Janssens, N. Buys, Estimating genetics of body dimensions and activity levels in pigs using automated pose estimation, *Sci. Rep.* 12 (2022) 15384.
 - [56] E. Huang, A. Mao, H. Gan, M. Camila Ceballos, T. D. Parsons, Y. Xue, K. Liu, Center clustering network improves piglet counting under occlusion, *Comput. Electron. Agric.* 189 (2021) 106417.
 - [57] J. Jia, S. Zhang, Q. Ruan, PCR: a large-scale benchmark for pig counting in real world, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2024.
 - [58] J. Kim, Y. Suh, J. Lee, H. Chae, H. Ahn, Y. Chung, D. Park, EmbeddedPigCount: pig counting with video object detection and tracking on an embedded board, *Sensors* (2022).
 - [59] S. Liu, C. Zhao, H. Zhang, Q. Li, S. Li, Y. Chen, R. Gao, R. Wang, X. Li, ICNet: a dual-branch instance segmentation network for high-precision pig counting, *AGRIC-Basel* (2024).
 - [60] Y. Luo, C. Yang, E. Lv, A. Yang, F. Meng, H. Luo, A lightweight model for automatic pig counting in intensive piggeries using a green inspection robot and image segmentation method, *Smart Agric. Technol.* (2025) 101115.
 - [61] M. Tian, H. Guo, H. Chen, Q. Wang, C. Long, Y. Ma, Automated pig counting using deep learning, *Comput. Electron. Agric.* 163 (2019) 104840.
 - [62] Y. Wang, D. Yang, H. Chen, L. Wang, Y. Gao, Pig counting algorithm based on improved YOLOv5n model with multiscale and fewer number of parameters, *Animals* 13 (2023) 3411.
 - [63] Y. Huang, D. Xiao, J. Liu, Z. Tan, K. Liu, M. Chen, An improved pig counting algorithm based on YOLOv5 and DeepSORT model, *Sensors* (2023).
 - [64] W. Feng, K. Wang, S. Zhou, An efficient neural network for pig counting and localization by density map estimation, *IEEE Access.* 11 (2023) 81079–81091.
 - [65] H. Ahn, S. Son, H. Kim, S. Lee, Y. Chung, D. Park, EnsemblePigDet: ensemble deep learning for accurate pig detection, *Appl. Sci.-Basel* (2021).
 - [66] F. Chen, X. Liang, L. Chen, B. Liu, Y. Lan, Novel method for real-time detection and tracking of pig body and its different parts, *Int. J. Agric. Biol. Eng.* (2020).
 - [67] Z. Hu, H. Yang, T. Lou, Dual attention-guided feature pyramid network for instance segmentation of group pigs, *Comput. Electron. Agric.* 186 (2021) 106140.
 - [68] E. Huang, Z. He, A. Mao, M. Ceballos, T. Parsons, K. Liu, A semi-supervised generative adversarial network for amodal instance segmentation of piglets in farrowing pens, *Comput. Electron. In Agric.* (2023).
 - [69] Z. Huang, H. Li, X. Xiong, X. He, R. Song, M. Huang, End-to-end individual pig detection based on transfer learning, in: 2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence, PRAI 2023, 2023.
 - [70] D. Liu, A. Parmiggiani, E. Psota, R. Fitzgerald, T. Norton, Where's your head at? Detecting the orientation and position of pigs with rotated bounding boxes, *Comput. Electron. Agric.* 212 (2023) 108099.
 - [71] M. Mattina, A. Benzinou, K. Nasreddine, F. Richard, An efficient anchor-free method for pig detection, *IET Image Process.* 17 (2023) 613–626.
 - [72] A. Parmiggiani, D. Liu, E. Psota, R. Fitzgerald, T. Norton, Don't get lost in the crowd Graph convolutional network for online animal tracking in dense groups, *Comput. Electron. Agric.* 212 (2023) 108038.
 - [73] J. Sha, G. Zeng, Z. Xu, Y. Yang, A light-weight and accurate pig detection method based on complex scenes, *Multimed. Tools Appl.* (2023).
 - [74] D. Sun, L. Zhang, J. Wang, X. Liu, Z. Wang, Z. Hui, J. Wang, Efficient and accurate detection of herd pigs based on Ghost-YOLOv7-SIoU, *Neural Comput. Appl.* 36 (2024) 2339–2352.
 - [75] M. Tian, F. Zhao, N. Chen, D. Zhang, J. Li, H. Li, Application of convolutional neural network in livestock target detection of IoT images, in: 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology, AINIT 2024, 2024.
 - [76] J.-H. Witte, J. Gerberding, C. Melching, J. M. Gómez, Evaluation of deep learning instance segmentation models for pig precision livestock farming, *Bus. Inf. Syst.* (2021) 209–220.
 - [77] K. Lei, C. Zong, T. Yang, S. Peng, P. Zhu, H. Wang, G. Teng, X. Du, Detection and analysis of sow targets based on image vision, *Agric.-Basel* (2022).
 - [78] L. Zhang, H. Gray, X. Ye, L. Collins, N. Allinson, Automatic individual pig detection and tracking in pig farms, *Sensors* (2019).
 - [79] Q. Du, G. Lu, L. Tian, Z. Sun, Z. Wang, Pig face recognition based on multi supervising forms, in: 2023 4th International Conference on Intelligent Computing and Human-Computer Interaction, ICHCI 2023, 2023.
 - [80] R. Ma, H. Ali, S. Chung, S. Kim, H. Kim, A lightweight pig face recognition method based on automatic detection and knowledge distillation, *APPL. SCI.-BASEL* (2024).
 - [81] W. Shigang, W. Jian, C. Meimei, W. Jinyang, A pig face recognition method for distinguishing features, in: 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2021, pp. 972–976, 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC).
 - [82] R. Wang, R. Gao, Q. Li, J. Dong, Pig face recognition based on metric learning by combining a residual network and attention mechanism, *Agric.-Basel* (2023).
 - [83] J. Guo, Y. Kong, L. Lin, L. Xu, D. Feng, L. Cao, J. Chen, J. Ye, S. Ye, Z. Yao, Y. Liu, T. Liu, S. Liu, Lightweight network based on Fourth order Runge-Kutta scheme and Hybrid Attention Module for pig face recognition, *Comput. Electron. Agric.* (2024).
 - [84] M. Mora, M. Piles, I. David, G. Rosa, Integrating computer vision algorithms and RFID system for identification and tracking of group-housed animals: an example with pigs, *J. Anim. Sci.* 102 (2024).
 - [85] Y. Pan, Y. Zhang, X. Wang, X. X. Gao, Z. Hou, Low-cost livestock sorting information management system based on deep learning, *Artif. Intell. Agric.* 9 (2023) 110–126.
 - [86] M. Wutke, D. Debiasi, S. Tomar, J. Probst, N. Kemper, K. Gevers, M. Lieboldt, I. Traulsen, Multistage pig identification using a sequential ear tag detection pipeline, *Sci. Reports* (2025).
 - [87] Z. Yin, M. Peng, Z. Guo, Y. Zhao, Y. Li, W. Zhang, F. Li, X. Guo, Lightweight pig face feature learning evaluation and application based on attention mechanism and two-stage transfer learning, *Agric.-Basel* (2024).
 - [88] F. Zhao, M. Tian, S. Hu, J. Liang, L. Zhou, H. Li, Livestock recognition and identification with deep convolutional neural networks: a case study of pigs, in: *Proceedings of the 2023 9th International Conference on Computing and Artificial Intelligence ICCAI '23*, New York, NY, USA, Association for Computing Machinery, 2023, pp. 185–194.
 - [89] S. Devi, J. Doley, J. Bharati, N. Mohan, V. Gupta, Analysis of pig posture detection in group-housed pigs using deep learning-based mask scoring instance segmentation, *Anim. Sci. J.* 95 (2024).
 - [90] S. Küster, P. Nolte, C. Meckbach, B. Stock, I. Traulsen, Automatic behavior and posture detection of sows in loose farrowing pens based on 2D-video images, *Front. Anim. Sci.* (2021).
 - [91] D. Liu, A. Parmiggiani, T. Norton, MT-SRNet: a transferable multi-task super-resolution network for pig keypoint detection, segmentation, and posture estimation, *Comput. Electron. Agric.* 237 (2025) 110533.
 - [92] M. S. Lyne, K. S. Henriksen, T. B. Moeslund, Autonomous monitoring of finishing pigs using side-view cameras and deep learning, in: *Proceedings of SPIE - The International Society for Optical Engineering*, 2022.
 - [93] M. Oczak, J.-L. Rault, S. Truong, O. Schmitt, Skeleton-based image feature extraction for automated behavioral analysis in human-animal relationship tests, *Appl. Anim. Behav. Sci.* 277 (2024) 106347.
 - [94] J.-J. Qin, L. Mao, J. Wang, Y. Li, L.-L. Wang, J. Li, A method for estimating pig pose in a breeding scenario, in: *Proceedings of SPIE - The International Society for Optical Engineering*, 2024.
 - [95] S. Tu, W. Yuan, Y. Liang, F. Wang, H. Wan, Automatic detection and segmentation for group-housed pigs based on PigMS R-CNN, *Sensors* (2021).
 - [96] W. Shi, X. Wang, X. Li, Y. Fu, X. Liu, H. Wang, YOLOv5DA: an improved YOLOv5 model for posture detection of grouped pigs, *Appl. Sci.* 14 (2024) 10104.
 - [97] M. Wang, W. Li, W. Zhang, G. Li, K. Dong, L. Zheng, Y. He, Y. Chen, Single pig pose estimation using cross-stage stacked hourglass network, in: *Proceedings of SPIE - The International Society for Optical Engineering*, 2023.
 - [98] M. Mattina, A. Benzinou, K. Nasreddine, F. Richard, An efficient center-based method for real-time pig posture recognition and tracking, *Appl. Intell.* 54 (2024) 5183–5196.
 - [99] A. Compte, Y. Yan, X. Cortés, S. Escalera, J. C. S. Jacques-Junior, Housed pig identification and tracking for precision livestock farming, *Expert Syst. Appl.* (2025) 128466.
 - [100] S. N. Adu Tagoe, M. Elmir, N. Amanquah, Monitoring of animal movement using computer vision, in: 2024 IEEE 9th International Conference on Adaptive Science and Technology (ICAST) 293, 2024, pp. 1–6, 2024 IEEE 9th International Conference on Adaptive Science and Technology (ICAST).
 - [101] H. Gan, M. Ou, F. Zhao, C. Xu, S. Li, C. Chen, Y. Xue, Automated piglet tracking using a single convolutional neural network, *Biosyst. Eng.* 205 (2021) 48–63.
 - [102] Lu J, Chen Z, Li X, Fu Y, Xiong X, Liu X and Wang H, 2024. ORP-Byte: a multi-object tracking method of pigs that combines Oriented RepPoints and improved ByteTrack computers and electronics in agriculture.
 - [103] T. Liu, D. Jie, J. Zhuang, D. Zhang, J. He, SDGTrack: a multi-target tracking method for pigs in multiple farming scenarios, *Animals* (2025).
 - [104] C. B. Dev Narayan, F. Rahman, M. Ullah, F. A. Cheikh, A. S. Imran, C. Coello, Ø. Nordbø, G. Santhosh Kumar, M. S. Nair, Tracking-by-self detection: a self-supervised framework for multiple animal tracking, *IFIP Adv. Inf. Commun. Technol.* (2023).
 - [105] A. Odo, N. McLaughlin, I. Kyriazakis, Re-identification for long-term tracking and management of health and welfare challenges in pigs, *Biosyst. Eng.* 251 (2025) 89–100.
 - [106] S. Gao, J. Gong, P. Yang, C. Liang, L. Huang, A stable long-term tracking method for group-housed pigs, in: H. Lu, W. Ouyang, H. Huang, J. Lu, R. Liu, J. Dong,

- M. Xu (Eds.), Image and Graphics, Springer Nature Switzerland, Cham, 2023, pp. 238–249.
- [107] S. Tu, Q. Zeng, H. Liu, Y. Liang, X. Liu, L. Huang, Z. Huang, MaskDis R-CNN: an instance segmentation algorithm with adversarial network for herd pigs, *IET Image Process.* 17 (2023) 3488–3499.
- [108] M. Wang, M. L. V. Larsen, D. Liu, J. F. M. Winters, J.-L. Rault, T. Norton, Towards re-identification for long-term tracking of group housed pigs, *Biosyst. Eng.* (2022).
- [109] L. Liu, J. Chen, Q. Ding, R. Zhao, M. Shen, L. Liu, Detection and analysis of sow nursing behavior based on the number and location of piglets outside the suckling area using YOLOv5s, *Comput. Electron. Agric.* (2025).
- [110] C. Song, X. Rao, Behaviors detection of pregnant sows based on deep learning, in: *ASABE 2018 Annual International Meeting*, 2018.
- [111] Y. Tsai, Y. Huang, E. Lin, S. Lai, X. Hong, J. Tsai, C. Chiang, Y. Kuo, Monitoring the lactation-related behaviors of sows and their piglets in farrowing crates using deep learning, *Front. Anim. Sci.* (2024).
- [112] A. Yang, H. Huang, X. Zhu, X. Wang, P. Chen, S. Li, Y. Xue, Automatic recognition of sow nursing behaviour using deep learning-based segmentation and spatial and temporal features, *Biosyst. Eng.* (2018).
- [113] I. Almadani, M. Abuhusseini, A. L. Robinson, YOLOv8-based estimation of estrus in sows through reproductive organ swelling analysis using a single camera, *Digital* (2024).
- [114] J.-H. Witte, J. Gerberding, C. Lensches, I. Traulsen, Using deep learning for automated birth detection during farrowing lecture notes in informatics (LNI), in: *Proceedings - Series of the Gesellschaft für Informatik (GI)*, 2022.
- [115] Y. Duan, Y. Yang, Y. Cao, X. Wang, R. Cao, G. Hu, Z. Liu, Integrated convolution and attention enhancement-you only look once: a lightweight model for false estrus and estrus detection in sows using small-target vulva detection, *ANIMALS* (2025).
- [116] M. Oczak, F. Bayer, S. Vetter, K. Maschat, J. Baumgartner, Comparison of the accelerated monitoring of the sow activity in farrowing pens using video and accelerometer data, *Comput. Electron. Agric.* 192 (2022) 106517.
- [117] M. Oczak, K. Maschat, J. Baumgartner, Implementation of computer-vision-based farrowing prediction in pens with temporary sow confinement, *Vet. Sci.* (2023) 109.
- [118] H. Gan, F. Menegon, A. Sun, A. Scollo, Q. Jiang, Y. Xue, T. Norton, Peeking into the unseen: occlusion-resistant segmentation for preweaning piglets under crushing events, *Comput. Electron. Agric.* 219 (2024) 108683.
- [119] H. Gan, X. Wu, F. Menegon, A. Scollo, A. Sun, T. Norton, Y. Xue, Occlusion-robust detection of sow-induced piglet crushing incidents using spatial and motion reasoning, *Comput. Electron. Agric.* 231 (2025) 109961.
- [120] E. Huang, A. Mao, J. Hou, Y. Wu, W. Xu, M. Camila Ceballos, T. D. Parsons, K. Liu, Occlusion-resistant instance segmentation of piglets in farrowing pens using center clustering network, *Comput. Electron. Agric.* 210 (2023) 107950.
- [121] Anon Deep-learning-based behavioral time budgets for sows with high and low piglet mortality rates.
- [122] H. Xue, J. Chen, Q. Ding, Y. Sun, M. Shen, L. Liu, X. Chen, J. Zhou, Automatic detection of sow posture and estrus based on convolutional neural network, *Front. Phys.* (2022).
- [123] Q. Ding, J. Chen, M. Shen, L. Liu, Activity detection of suckling piglets based on motion area analysis using frame differences in combination with convolution neural network, *Comput. Electron. Agric.* 202 (2022) 107423.
- [124] L. van der Zande, O. Guzhva, S. Parois, I. van de Leemput, J. Bolhuis, T. Rodenburg, Estimation of resilience parameters following LPS injection based on activity measured with computer vision, *Front. Anim. Sci.* (2022).
- [125] R. L. Jeon, J. M. Peschel, B. C. Ramirez, J. D. Stock, K. J. Stalder, Deep learning based landmark detection for measuring hock and knee angles in sows, *Trans. Anim. Sci.* 8 (2024) txad033.
- [126] A. Nasirahmadi, B. Sturm, S. Edwards, K. Jeppsson, A. Olsson, S. Müller, O. Hensel, Deep learning and machine vision approaches for posture detection of individual pigs, *Sensors* (2019).
- [127] G. de Paula T. M. C., V. de Sousa R., M. P. Sarmiento, T. Kramer, J. de Souza Sardinha E., L. Sabei, J. S. Machado, M. Villoti, A. J. Zanella, Deep learning pose detection model for sow locomotion, *Sci. Rep.* 14 (2024) 16401.
- [128] M. Rieker, A. Klein, F. Adion, C. Hoffmann, E. Gallmann, Automatically detecting pig position and posture by 2D camera imaging and deep learning, *Comput. Electron. Agric.* 174 (2020) 105391.
- [129] S. Bery, T. Brown-Brandl, B. Jones, G. Rohrer, S. Sharma, Determining the presence and size of shoulder lesions in sows using computer vision, *Animals* (2024).
- [130] H. Gan, M. Ou, C. Li, X. Wang, J. Guo, A. Mao, M. Camila Ceballos, T. D. Parsons, K. Liu, Y. Xue, Automated detection and analysis of piglet suckling behaviour using high-accuracy amodal instance segmentation, *Comput. Electron. Agric.* (2022).
- [131] T. Van De Putte, C. Van Kerschaver, M. Hostens, J. Degroote, Investigating the development of piglet feeding behaviour during the immediate postweaning phase using computer vision, *Animal* (2025).
- [132] Y. Wang, G. Sun, X. Seng, H. Zheng, H. Zhang, T. Liu, Deep learning method for rapidly estimating pig body size, *Anim. Prod. Sci.* 63 (2023) 909–923.
- [133] X. Dong, C. Zhang, P. Wang, D. Chen, G. Tu, S. Zhao, T. Xiang, A novel dual-network approach for real-time liveweight estimation in precision livestock management, *Adv. Sci.* (2025).
- [134] J. Liu, D. Xiao, Y. Liu, Y. Huang, A pig mass estimation model based on deep learning without constraint, *Animals* (2023).
- [135] Z. Liu, W. Peng, C. Wang, G. Li, B. Ji, D. Casey, N. Lu, Y. Zhao, Automatic swine body measurements via two-dimensional imaging system and correlation analysis with conformation scoring, *Smart Agric. Technol.* (2025).
- [136] S. Wan, H. Fang, X. Wang, Monocular visual pig weight estimation method based on the EfficientVit-C model, *Agric. (Switz.)* (2024).
- [137] A. C. R. da Cunha, R. C. Antunes, W. G. da Costa, G. F. Rebouças, C. D. S. Leite, A. S. do Carmo, Body weight prediction in crossbred pigs from digital images using computer vision, *Livest. Sci.* 282 (2024) 105433.
- [138] J. Fraser, H. Aricibasi, D. Tulpan, R. Bergeron, A computer vision image differential approach for automatic detection of aggressive behavior in pigs using deep learning, *J. Anim. Sci.* (2023).
- [139] M. Bonneau, N. Pouillet, D. Beramice, L. Dantec, L. Canario, J. Gouridine, Behavior comparison during chronic heat stress in large white and creole pigs using image-analysis, *Front. Anim. Sci.* (2021).
- [140] A. Odo, N. McLaughlin, I. Kyriazakis, Automated monitoring of ear biting in pigs by tracking individuals and events, in: *Proceedings - 2024 IEEE Winter Conference on Applications of Computer Vision, WACV 2024*, 2024.
- [141] C. Chen, W. Zhu, M. Oczak, K. Maschat, J. Baumgartner, M. L. V. Larsen, T. Norton, A computer vision approach for recognition of the engagement of pigs with different enrichment objects, *Comput. Electron. Agric.* (2020).
- [142] H. Gan, M. Ou, E. Huang, C. Xu, S. Li, J. Li, K. Liu, Y. Xue, Automated detection and analysis of social behaviors among preweaning piglets using key point-based spatial and temporal features, *Comput. Electron. Agric.* 188 (2021) 106357.
- [143] H. Gan, C. Xu, W. Hou, J. Guo, K. Liu, Y. Xue, Spatiotemporal graph convolutional network for automated detection and analysis of social behaviours among preweaning piglets, *Biosyst. Eng.* (2022).
- [144] J. Han, J. Siegfried, D. Colbry, R. Lesiyon, A. Bosgraaf, C. Chen, T. Norton, J. P. Steibel, Evaluation of computer vision for detecting agonistic behavior of pigs in a single-space feeding stall through blocked cross-validation strategies, *Comput. Electron. Agric.* 204 (2023) 107520.
- [145] F. Hakansson, D. B. Jensen, Parameter testing and systematic performance assessment of machine-learning models for video-based classification of damaging social behaviour in groups of pigs, in: *Precision Livestock Farming 2022 - Papers Presented at the 10th European Conference on Precision Livestock Farming, ECPLF 2022*, 2022.
- [146] F. Hakansson, D. Jensen, Automatic monitoring and detection of tail-biting behavior in groups of pigs using video-based deep learning methods, *Front. Vet. Sci.* (2023).
- [147] H. Ji, G. Teng, J. Yu, Y. Wen, H. Deng, Y. Zhuang, Efficient aggressive behavior recognition of pigs based on temporal shift module, *Animals* (2023).
- [148] P. Kasani, S. Oh, Y. Choi, S. Ha, H. Jun, K. Park, H. Ko, J. Kim, J. Choi, E. Cho, J. Kim, A computer vision-based approach for behavior recognition of gestating sows fed different fiber levels during high ambient temperature, *J. Anim. Sci. And Technol.* (2021).
- [149] H. Kim, Y. Kim, F. Devira, M. Yi, Application of anomaly detection to identifying aggressive pig behaviors using reconstruction loss inversion, *Agriengineering* (2024).
- [150] M. L. V. Larsen, M. Wang, S. Willems, D. Liu, T. Norton, Automatic detection of locomotor play in young pigs: a proof of concept, *Biosyst. Eng.* (2023).
- [151] D. Liu, M. Oczak, K. Maschat, J. Baumgartner, B. Pletzer, D. He, T. Norton, A computer vision-based method for spatial-temporal action recognition of tail-biting behaviour in group-housed pigs, *Biosyst. Eng.* (2020).
- [152] M. Oczak, F. Bayer, S. G. Vetter, K. Maschat, J. Baumgartner, Where is the sow's nose: RetinaNet object detector as a basis for monitoring the use of rack with nest-building material, *Front. Anim. Sci.* (2022).
- [153] C. Qin, Y. Li, H. Jia, Intelligent livestock farming: monitoring pig behavior with enhanced YOLOv5 for spatial temporal feature fusion, *Int. J. Intell. Comput. Cybern.* 18 (2025) 444–461.
- [154] G. Taiwo, S. Vadera, A. Alameer, Vision transformers for automated detection of pig interactions in groups, *Smart Agric. Technol.* (2025) 100774.
- [155] J. Wei, X. Tang, J. Liu, Z. Zhang, Detection of pig movement and aggression using deep learning approaches, *Animals* (2023).
- [156] Q. Yang, M. Chen, D. Xiao, S. Huang, X. Hui, Long-term video activity monitoring and anomaly alerting of group-housed pigs, *Comput. Electron. Agric.* 224 (2024) 109205.
- [157] R. Yang, Z. Chen, H. Xu, M. Shen, P. Li, T. Norton, M. Lu, Recognizing the rooting action of prepartum sow in free-farrowing pen using computer vision, *Comput. Electron. Agric.* 213 (2023) 108167.
- [158] M. Hansen, E. Baxter, K. Rutherford, A. Futro, M. Smith, L. Smith, Towards facial expression recognition for on-farm welfare assessment in pigs, *Agric.-Basel* (2021).
- [159] M. Wutke, C. Lensches, U. Hartmann, I. Traulsen, Towards automatic farrowing monitoring—a noisy student approach for improving detection performance of newborn piglets, *PLoS. One* 19 (2024) e0310818.
- [160] Marsot M, Mei J, Shan X, Ye L, Feng P, Yan X, Li C and Zhao Y, 2020. An adaptive pig face recognition approach using Convolutional Neural Networks Computers and Electronics in Agriculture.
- [161] C. Chen, W. Zhu, J. Steibel, J. Siegfried, J. Han, T. Norton, Recognition of feeding behaviour of pigs and determination of feeding time of each pig by a video-based deep learning method, *Comput. Electron. Agric.* 176 (2020) 105642.
- [162] M. Oczak, F. Bayer, K. Maschat, J. Baumgartner, Computer vision for monitoring hay rack use behavior in sows, in: *Precision Livestock Farming 2022 - Papers Presented at the 10th European Conference on Precision Livestock Farming, ECPLF 2022*, 2022.
- [163] M. Bonneau, B. Benet, Y. Labrune, J. Bailly, E. Ricard, L. Canario, Predicting sow postures from video images: comparison of convolutional neural networks and segmentation combined with support vector machines under various training and testing setups, *Biosyst. Eng.* (2021).

- [164] J. Chen, L. Liu, P. Li, W. Yao, M. Shen, L. Liu, Resting posture recognition method for suckling piglets based on piglet posture recognition (PPR)—you only look once, *Agriculture* 15 (2025) 230.
- [165] O. Girardie, M. Bonneau, Y. Billon, J. Bailly, I. David, L. Canario, Analysis of image-based sow activity patterns reveals several associations with piglet survival and early growth, *Front. Vet. Sci.* (2023).
- [166] A. Scaillierez, T. Garcia-Faria, H. Broers, S. Koning, R. van der Tol, E. Bokkers, I. Boumans, Determining the posture and location of pigs using an object detection model under different lighting conditions, *Transl. Anim. Sci.* (2024).
- [167] J.-H. Witte, P. Heseke, J. Probst, N. Kemper, I. Traulsen, J. M. Gómez, Tail posture as a predictor of tail biting in pigs: a camera-based monitoring system, in: 11th European Conference on Precision Livestock Farming, 2024.
- [168] J.-H. Witte, P. Heseke, J. Probst, I. Traulsen, N. Kemper, J. M. Gómez, Image-based tail posture monitoring of pigs, in: *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2024.
- [169] Z. Wang, S. Zhou, P. Yin, A. Xu, J. Ye, GANPose: pose estimation of grouped pigs using a generative adversarial network, *Comput. Electron. Agric.* 212 (2023) 108119.
- [170] S. Tu, Y. Cao, Y. Liang, Z. Zeng, H. Ou, J. Du, W. Chen, Tracking and automatic behavioral analysis of group-housed pigs based on YOLOX+BoT-SORT-slim, *Smart Agric. Technol.* (2024).
- [171] H. Gan, M. Ou, F. Zhao, C. Xu, S. Li, C. Chen, Y. Xue, Automated piglet tracking using a single convolutional neural network, *Biosyst. Eng.* 205 (2021) 48–63.
- [172] C. Chen, W. Zhu, J. Steibel, J. Siegfors, J. Han, T. Norton, Classification of drinking and drinker-playing in pigs by a video-based deep learning method, *Biosyst. Eng.* 196 (2020) 1–14.
- [173] Z. Liu, J. Liu, Y. Zhou, Z. Wang, A method for pig behavior target detection based on the lightweight RTMDet model, in: 2024 International Symposium on Digital Home (ISDH), 2024, pp. 25–30, 2024 International Symposium on Digital Home (ISDH).
- [174] W. Shi, X. Wang, X. Li, Y. Fu, X. Liu, H. Wang, YOLOv5DA: an improved YOLOv5 model for posture detection of grouped pigs, *Appl. Sci.* 14 (2024) 10104.
- [175] J. Chen, L. Liu, P. Li, W. Yao, M. Shen, L. Liu, Resting posture recognition method for suckling piglets based on piglet posture recognition (PPR)—you only look once, *Agric. (Switz.)* (2025).
- [176] A. C. R. da Cunha, R. C. Antunes, W. G. da Costa, G. F. Rebouças, C. D. S. Leite, A. S. do Carmo, Body weight prediction in crossbred pigs from digital images using computer vision, *Livest. Sci.* 282 (2024) 105433.
- [177] Y. Wang, G. Sun, X. Seng, H. Zheng, H. Zhang, T. Liu, Deep learning method for rapidly estimating pig body size, *Anim. Prod. Sci.* 63 (2023) 909–923.
- [178] R.-W. Bello, D. A. Olubummo, Computer vision-based precision livestock farming: an overview of the challenges and opportunities, *SSRN J.* (2024).
- [179] W. Shigang, W. Jian, C. Meimei, W. Jinyang, A pig face recognition method for distinguishing features, in: *Proceedings of IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers, IPEC 2021*, 2021.
- [180] S. Wang, H. Jiang, Y. Qiao, S. Jiang, H. Lin, Q. Sun, The research progress of vision-based artificial intelligence in smart Pig farming, *Sensors* 22 (2022).
- [181] H.-R. Nasser, C. Kasper, Unlocking the potential of computer vision in precision pig farming: a call for a collaborative data and AI models platform, *Anim. Fron.* (2026) vfag003.
- [182] G. L. Menezes, G. Mazon, R. E. P. Ferreira, V. E. Cabrera, J. R. R. Dorea, Artificial intelligence for livestock: a narrative review of the applications of computer vision systems and large language models for animal farming, *Anim. Front.* 14 (2024) 42–53.