Expert Systems With Applications 289 (2025) 128330

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Animal behavior analysis methods using deep learning: A survey

Edoardo Fazzari [a,b,c,*], Donato Romano [a,b], Fabrizio Falchi [a,c], Cesare Stefanini [a,b]

[a] *The BioRobotics Institute, Sant'Anna School of Advanced Studies, Viale Rinaldo Piaggio, Pontedera, 56025, Italy*
[b] *Department of Excellence in Robotics and AI, Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà, Pisa, 56127, Italy*
[c] *Institute of Information Science and Technologies, National Research Council of Italy, via G. Moruzzi, Pisa, 56124, Italy*

ARTICLE INFO

Keywords:
Animal behavior
Deep learning
Pose estimation
Object detection
Bio-acoustics
Machine learning

ABSTRACT

Animal behavior serves as a reliable indicator of the adaptation of organisms to their environment and their overall well-being. Through rigorous observation of animal actions and interactions, researchers and observers can glean valuable insights into diverse facets of their lives, encompassing health, social dynamics, ecological relationships, and neuroethological dimensions. Although state-of-the-art deep learning models have demonstrated remarkable accuracy in classifying various forms of animal data, their adoption in animal behavior studies remains limited. This survey article endeavors to comprehensively explore deep learning architectures and strategies applied to the identification of animal behavior, spanning auditory, visual, and audiovisual methodologies. The survey categorizes techniques into pose estimation-based and non-pose estimation-based methods, analyzing their applications, effectiveness, and limitations. Furthermore, the manuscript scrutinizes extant animal behavior datasets, offering a detailed examination of the principal challenges confronting this research domain. The article culminates in a comprehensive discussion of key research directions within deep learning that hold potential for advancing the field of animal behavior studies.

1. Introduction

The study of animal behavior involves the observation, description, and comprehension of how animals engage with one another and their surroundings. Presently, the landscape of animal behavior research is undergoing rapid evolution, propelled by the continual introduction of innovative experimental methodologies and the advancement of sophisticated behavior detection systems (Wang, Du, Wang, Hu, & Zhao, 2021b). This progression holds particular significance in advancing our understanding of neuroethological aspects, exemplified by the utilization of mice in exploring diseases like Alzheimer's (Pedersen et al., 2006), and in refining animal welfare practices within agriculture (Mishra & Sharma, 2023). The impetus behind this surge in progress is the integration of cutting-edge technologies, with deep learning standing out as a transformative force that reshapes the approaches researchers employ to investigate and interpret animal behaviors (Brown & de Bivort, 2017).

Deep learning has emerged as a pivotal tool in the exploration of animal behavior. This advanced branch of artificial intelligence empowers computers to autonomously discern patterns and features from extensive datasets, improving upon earlier methods based on classical Machine Learning, as discussed in Valletta, Torney, Kings, Thornton, and Madden (2017). As researchers amass increasingly intricate datasets through state-of-the-art monitoring technologies, including high-resolution cameras, GPS tracking devices, and sensors, the capability of deep learning algorithms to extract meaningful insights becomes indispensable (Benaissa et al., 2023; Koger et al., 2023). This not only expedites the analysis process but also reveals nuanced aspects of animal behavior that were previously challenging to decipher. Furthermore, deep learning plays a crucial role in the development of sophisticated behavior detection systems. These systems can automatically recognize and classify various behaviors, allowing researchers to redirect their focus from laborious manual data annotation to the interpretation of results (Arablouei et al., 2023a). This acceleration in data processing and behavior recognition enhances the scalability and efficiency of animal behavior studies, ushering in a new era of discovery and understanding in this dynamic field.

This survey article makes a threefold contribution to the current understanding of the study of animal behavior through deep learning:

- We provide a thorough examination of existing technologies and algorithms employed in the analysis of animal behavior. This entails

---

* Corresponding author.
E-mail addresses: edoardo.fazzari@santannapisa.it, edoardo.fazzari@isti.cnr.it (E. Fazzari), donato.romano@santannapisa.it (D. Romano), fabrizio.falchi@cnr.it (F. Falchi), cesare.stefanini@santannapisa.it (C. Stefanini).

https://doi.org/10.1016/j.eswa.2025.128330
Received 11 December 2024; Received in revised form 30 April 2025; Accepted 23 May 2025

a detailed exploration of methodologies and approaches currently prevalent in this domain, providing readers with a nuanced understanding of the technological landscape.

- We compile and present a comprehensive list of publicly available datasets relevant to the research field. This compilation serves as a valuable resource for researchers and practitioners, facilitating access to essential data for furthering investigations into animal behaviors through data-driven methodologies.
- We engage in a substantive discussion regarding potential directions for the evolution of the field. Emphasizing the integration of deep learning techniques, our discourse aims to enhance the quality of existing technologies, thereby advancing the understanding of animal behaviors. This forward-looking analysis provides insights into potential avenues for improvement and innovation.

To the best of our knowledge, this survey is the only examination of the topic to date. The comprehensive overview, dataset compilation, and forward-looking discussions collectively contribute to a nuanced understanding of current advancements and lay the groundwork for future developments in the application of deep learning to the study of animal behavior.

2. Motivation and problem statement

In this section, we expound upon the foundational motivations that underscore the study of animal behavior through deep learning, articulating the diverse advantages and objectives inherent to this specialized research domain. Concurrently, we meticulously scrutinize the principal limitations that characterize this field, recognizing the nuanced challenges that arise from variations in research setups. Our intention is to furnish a comprehensive guide for prospective researchers, endowing them with a thorough understanding of potential impediments prior to initiating investigations within this domain. Concurrently, we underscore the myriad opportunities inherent in the study of animal behavior, fostering an appreciation for the intricate dimensions of this research frontier.

2.1. Limitations in studying animal behaviors

The exploration of animal behavior is confronted by a multitude of challenges that intricately shape the effectiveness and practicality of its applications. These challenges permeate the methodologies employed for data acquisition, the intricacies of data analysis (both in terms of location and computational demands), and the nuanced process of data annotation.

An integral aspect of animal behavior studies is the utilization of sensors for data collection. However, the attachment of sensors to animals introduces a unique set of challenges. Notably, there is a risk of inducing stress responses and altering normal behaviors (Zhang, Li, Huang, & Chen, 2020). This necessitates a profound reflection on the authenticity of the collected data, urging researchers to question whether stress-induced behaviors accurately mirror natural patterns. Moreover, the task of differentiating genuine behavioral signals from background noise in sensor data adds complexity to interpretation, emphasizing the requirement for advanced algorithms capable of discerning meaningful patterns amidst the noise (Kavlak, Pastell, & Uimari, 2023). Beyond stress responses and noise challenges, sensor equipment grapples with limitations in battery life (Mekruksavanich, Jantawong, & Jitpattanakul, 2022), impacting the duration and scope of behavioral studies, especially in scenarios requiring continuous data collection over extended periods. Researchers are challenged to strike a balance between the need for comprehensive, continuous monitoring and the practical constraints imposed by limited battery capacities.

Transitioning into the domain of mobile devices introduces another layer of complexity to the challenges encountered in deep learning applications. The implementation of models on these devices confronts the persistent issue of storage limitations (Cao, Zhao, Liu, & Sun, 2020). Balancing robust object detection with efficient data compression becomes a paramount concern, with techniques like Quantized-CNN (Wu, Leng, Wang, Hu, & Cheng, 2016) attempting to address this challenge. However, the ongoing quest is to achieve this balance without compromising precision, a crucial consideration for the reliability of behavioral analyses. Furthermore, light-weight processing enables deployment of trained models on-board, reducing reducing latency and bandwidth requirements in harsh farming environments (Temenos et al., 2024).

A pivotal challenge arises in the realm of labeling and annotation, where economic and practical constraints hinder the tagging of large datasets for each animal (Bhattacharya & Shahnawaz, 2021). This bottleneck impedes the scalability of deep learning models, heavily reliant on labeled datasets for effective training. The impracticality of manual labeling raises fundamental concerns about the breadth and accuracy of behavioral datasets, impacting the reliability of subsequent analyses. Subjectivity compounds these challenges, influencing the accuracy and consistency of behavioral annotations. Visual inspection, often subjective, is limited in providing objective insights into complex animal behaviors (Bernardes, Lima, Guedes, da Silva, & Martins, 2021). Manual annotation, while traditional, is labor-intensive and susceptible to inter-annotator disagreements (Segalin et al., 2021; Zhou et al., 2022). The inherent subjectivity introduces variability, raising questions about the replicability and reliability of experiments (Dell et al., 2014; Hou et al., 2020).

Moreover, these challenges extend to innovative techniques, such as multi-view recordings, which hold promise for providing richer insights into animal behaviors (Jiang et al., 2021). However, challenges arise in correlating social behaviors from different perspectives due to the lack of correspondence across data sources. Effectively coordinating information from multiple viewpoints demands inventive approaches to ensure accuracy and reliability, representing a frontier where deep learning methodologies can contribute significantly.

A major challenge surfaces when comparing laboratory studies, where challenges often revolve around the subjectivity introduced by the controlled environment, with ethological studies conducted in the wild presenting a distinct set of challenges, notably in-field tracking (Marshall, Li, Wu, & Dunn, 2022). The diverse and unpredictable environments encountered in the wild introduce complexities not found in controlled laboratory settings. Bridging the gap between these two disciplines necessitates adaptable detection and tracking algorithms that seamlessly operate in both environments. Robust algorithms capable of handling varying animal sizes, changing appearances, clutter, occlusions, and unpredictable environments are vital for extracting meaningful insights (Haalck, Mangan, Webb, & Risse, 2020; Hou et al., 2020; Lauer et al., 2022). These challenges underscore the critical need for technological innovation that aligns with the demands of both controlled and wild settings.

2.2. Objectives in studying animal behaviors

Studying animal behavior offers manifold advantages, enriching our comprehension of the natural world and presenting practical applications across diverse domains, including neuroscience, pharmacology, medicine, agriculture, ecology, and robotics. Six key advantages of studying animal behavior are identified:

1. **Biodiversity conservation**: Understanding animal behavior is crucial for the conservation of biodiversity (Ditria et al., 2020; Hou et al., 2020; Nilsson et al., 2020; Pillai, Gupta, Sharma, & Bansal, 2023; Wijeyakulasuriya, Eisenhauer, Shaby, & Hanks, 2020). Knowledge of behaviors such as migration patterns, feeding habits, and reproductive strategies is essential for designing effective conservation strategies and protecting endangered species.
2. **Ecological understanding:** Animal behavior provides insights into the ecological dynamics of ecosystems (Akçay et al., 2020; Gotanda,

Farine, Kratochwil, Laskowski, & Montiglio, 2019). Behavioral studies help researchers understand how animals interact with their environment, including their roles in nutrient cycling, seed dispersal, and predator-prey relationships (Chen et al., 2020; Nasiri, Amirivojdan, Zhao, & Gan, 2023; Yamada, Shawe-Taylor, & Fountas, 2020).

3. **Human health and medicine:** Studying animal behavior can have implications for human health and medicine (Gnanasekar, Yanushkevich, Van den Hoogen, & Trang, 2022; Hart, 2011; Manduca et al., 2023; Shaw & Lahrman, 2023). For example, research on animal models helps in understanding certain diseases and developing potential treatments. Behavioral studies on animals also contribute to our understanding of the neurobiology and psychology that underlie human behavior (Coria-Avila et al., 2022; Mathis & Mathis, 2020; Saleh, Ahmed, Zaafan, Farouk, & Atia, 2023).
4. **Animal welfare and husbandry:** Knowledge of animal behavior is essential for promoting the welfare of domesticated animals and optimizing their husbandry practices (Bao & Xie, 2022; Jiang et al., 2021). Understanding how animals express natural behaviors can inform the design of environments that support their physical and psychological well-being Jiang, Rao, Zhang, and Shen (2020); Manoharan (2020); Tassinari et al. (2021).
5. **Pest control and agriculture:** Although right now it is very limited to pest identification, understanding the behavior of pest species can aid in the development of effective pest control strategies in agriculture (Coulibaly, Kamsu-Foguem, Kamissoko, & Traore, 2022; Júnior & Rieder, 2020; Mendoza et al., 2023). This knowledge helps farmers manage crop damage and reduce the need for harmful pesticides (Mankin, Hagstrum, Guo, Eliopoulos, & Njoroge, 2021; Teixeira, Ribeiro, Morais, Sousa, & Cunha, 2023).
6. **Understanding social dynamics:** Observing social behaviors in animals can provide insights into the principles governing social structures and interactions (Papaspyros et al., 2023; Xiao et al., 2023). This knowledge can have applications in fields such as sociology and psychology, contributing to our understanding of social dynamics in general (Alameer et al., 2022; Landgraf et al., 2021; Perez & Toler-Franklin, 2023).

In this context, deep learning plays a pivotal role and emerges as a major technology in advancing the field, opening new opportunities. Addressing the limitations discussed in the previous section, we will now elaborate on the advantages that deep learning and computational technologies bring to the study of animal behavior.

We previously emphasized the integral aspect of employing sensors for data collection in animal behavior studies, which may induce stress and high noise levels. The use of multiple and diverse sensors for data acquisition, coupled with advanced architectures incorporating fusion layers, has been shown to mitigate noise and enhance the precision of analysis (Mahmud, Zahid, Das, Muzammil, & Khan, 2021). However, attaching sensors directly to animals may introduce bias, prompting researchers in livestock health assessment and neuroscience to adopt computer vision. The ability of computer vision to provide real-time, non-invasive, and accurate animal-level information through the use of cameras has gained popularity (Oliveira, Pereira, Bresolin, Ferreira, & Dorea, 2021). Nevertheless, this approach is limited to setups within the camera frame, except for innovations like Haalck et al. (2020) moving camera that tracks animals, creating a dynamic map of their environment. In larger scenarios, such as meadows where cows graze, sensors remain preferable. Nevertheless, collecting and analyzing sensor data from mobile devices on animals proves challenging and time-consuming. To address this, Dang, Dang, Tran, and Chung (2022) introduced Long Range Area Network (LoRaWAN), where sensors attached to cows connect to gateways transmitting information to the cloud. This not only overcomes the limitations of computational power associated with mobile devices but also ensures continuous, real-time data analysis.

The efficacy of deep learning is contingent on annotated data, especially for supervised approaches. While manual annotation remains unavoidable, in tasks such as pose estimation and classification, labeling can be iterative. This involves annotating a small portion of the dataset, training a network, predicting on new images, correcting labels, and repeating this process multiple times. This iterative approach accelerates the labeling process, as demonstrated by Pereira et al. (2019). Another approach is to generate artificial labels (Li & Lee, 2023).

Table 1
Summary of the limitations and objectives in studying animal behaviors. DL stands for Deep Learning.

Limitations	Objectives
Sensor-induced stress	Biodiversity conservation
Battery life	Biodiversity preservation
Data noise	Ecological insight
Storage constraints	Health impact
Labeling economics	Welfare optimization
Subjective annotation	Pest management
Computational demands	Social dynamics
Ethical considerations	Non-invasive (DL related)
In-field tracking	Real-time applications (DL related)
Environmental unpredictability	

Table 1 succinctly encapsulates a consolidated overview of both primary limitations and advantages, providing a discerning reference for researchers navigating the sophisticated landscape of animal behavior studies.

2.3. Research questions

The survey aims to address the following research questions:

RQ1 *Which animal species are more considered and why?*
This research question seeks to investigate commonly explored and well-studied animal species in the context of behavior analysis. To address this question, the survey will explore existing literature and research to identify trends and biases in the selection of animal subjects. The objective is to understand why certain species are more frequently examined and to gain insights into potential research trends, contributing to a deeper understanding of established knowledge and guiding future studies in the field.

RQ2 *What deep learning methods have been used in the literature for animal behavior analysis?*
This research question focuses on summarizing and categorizing the existing deep learning methods employed in the literature for animal behavior analysis. The survey will review a wide range of studies to identify and classify the various deep learning techniques applied to analyze animal behavior. The objective is to analyze existing methodologies to identify trends, strengths, and limitations of current approaches in the field.

RQ3 *What are the deep learning strategies that are suitable and could enhance this task, but are not yet exploited?*
This research question looks forward, aiming to identify untapped potential in the application of deep learning to animal behavior analysis. The survey will involve a comprehensive review of the current literature to identify gaps or areas where deep learning strategies have not been extensively explored. This involves proposing novel applications of existing techniques or suggesting modifications to adapt deep learning methods for more effective analysis of animal behavior.

3. Method for literature survey

In this section, we clarify the methodology applied in our survey. Our approach involved a thorough systematic review to carefully select the pertinent studies considered in this article. Following this, we accurately analyzed the gathered information, employing statistical methods to derive meaningful insights.

Table 2
Features extracted from papers.

Reference:	Assigned identifier for each retrieved article.
Year:	Publication year of the article.
Country:	Geographical location where the authors are based, as required in Section 3.2.
Species:	The specific species under investigation in the article.
Pose estimation:	Indicates whether the methodology incorporates pose estimation (*True* or *False*).
Behavior analysis:	Indicates whether the analysis considers an association between extracted features and observed behaviors (*True* or *False*).
Feature methodology:	The approach employed to extract salient features in the study.
Behavior methodology:	The methodological framework used to correlate features with observed behaviors.
Authors' research field:	Indicates the research fields the authors mainly work on.

3.1. Search and selection strategies

This section delineates the methodologies employed for data collection and synthesis. Initially, data acquisition was conducted through systematic searches on academic repositories, including Google Scholar, IEEE Xplore, and the Springer Database. The formulated search queries were as follows:

animal behavior **AND** *deep learning*
(*insect* **OR** *wild*) **AND** *behavior* **AND** *deep learning*

The decision to employ distinct queries for insects and wild animals was necessitated by the observed paucity of literature in these categories relative to studies involving farm animals and neuroethology, commonly focused on mice. Thus, the formulation of specific queries was imperative to encompass a broader spectrum of animal species. Notably, while our search explicitly included “behavior”, we ensured that both American and British spellings (“behavior” and “behaviour”) were effectively considered. Our observations confirmed that academic indexing systems retrieve papers regardless of spelling variations, mitigating the need for separate queries for each variant.

Regarding deep learning, we opted to use “deep learning” as a general term rather than specifying particular architectures such as CNNs or RNNs. This choice was informed by the observation that papers discussing neural networks typically reference deep learning in some capacity. Indeed, several retrieved papers had titles mentioning CNNs or other architectures without explicitly stating “deep learning,” yet they addressed deep learning concepts in their introductions. Since referencing the broader field is common in academic writing, this approach ensured comprehensive coverage of relevant studies.

Lastly, the inclusion of “animal” in our search strategy required careful consideration. As indicated by our need for a distinct query for insects, many papers on insect studies did not explicitly mention the term “animal”. To address this, we incorporated (*insect* OR *wild*) into our search queries. The term “wild” was particularly valuable, as it serves as the root of “wildlife”, a term frequently associated with key research themes such as wildlife conservation, biodiversity, and ecological studies. This inclusion allowed us to capture a broader range of studies focusing on animals in non-domesticated settings, ensuring a more comprehensive literature retrieval.

Subsequently, the acquired data underwent systematic tabulation based on features explicated in Table 2. These features were derived from discerned patterns identified during a comprehensive analysis of the extant literature. Synthesizing the outcomes of this analysis, Sections 4 and 5 encapsulate the aggregated findings, summarizing the respective papers that expound upon solution methodologies grounded in pose estimation and those that do not.

Finally, a judicious filtering operation was executed to extract only those articles germane to the objectives of this survey, resulting in 161 articles. Each article within this subset underwent thorough examination, and pertinent references therein were scrutinized and subsequently incorporated into our survey to enrich its content.

The overall process is described via the PRIMA flowchart in Fig. 1.

3.2. Comprehensive science mapping analysis

3.2.1. Annual scientific production

In the process of retrieving articles, our attention was exclusively directed towards research publications spanning the temporal spectrum from 2020 to 2023. Fig. 2(a) elucidates this distribution through a histogram, illustrating the quantitative representation of papers across each respective year.

3.2.2. Scientific production based on animal considered

Fig. 2(b) illustrates the distribution of percentages pertaining to the various animal species under consideration in the selected articles. Evidently, a predominant emphasis is placed on research concerning livestock, notably focusing on cows and pigs, as well as studies involving mice, related mostly to neuroscience.

3.2.3. Research field of authors

Given that animal behavior analysis is an interdisciplinary and multidisciplinary field, it becomes imperative to comprehend the research background of individuals engaged in this domain. Despite the predominant focus on articles related to deep learning technologies, it is noteworthy that a considerable number of non-artificial intelligence practitioners are actively entering this field, as illustrated in Fig. 2(c). Interestingly, when combining "computer science" (encompassing computer engineering) and "artificial intelligence," they constitute only 18 % of the scholarly contributions. In contrast, bio-related fields, including biology, animal science, agriculture, veterinary, and ecology, collectively contribute 30 % to the research landscape. Noteworthy is the active participation of various engineering fields, even those with a mechanical-electrical background, in the exploration of animal behavior. Additionally, a compelling correlation is observed in the fields of neuroscience and psychology, where the majority of articles are dedicated to the study of mice.

4. Pose estimation-based methods

A potential first step in measuring behavior is to identify meaningful keypoints on the animal’s body to track the movements of specific body parts and quantify behavioral patterns.

Pose estimation, the process of identifying and locating the position and orientation of objects, is a fundamental technique widely used in the examination of animal behaviors alongside object detection, as discussed in Section 5.3. Originating from Human Pose Estimation (HPE), the evolution into Animal Pose Estimation (APE) was spearheaded by Mathis et al. (2018) through DeepLabCut and Pereira et al. (2019) via LEAP, subsequently evolving into SLEAP (Pereira et al., 2022). This section delves into an in-depth analysis of these two methodologies juxtaposed with emerging trends within the field of research. Given the primary focus of our survey on animal behavior analysis, subsequent to the introduction of these predominant approaches, we elucidate the utilization of pose estimation outputs for behavior analysis and classification. For a more comprehensive understanding of animal pose estimation, we recommend perusing the survey conducted by Jiang, Lee, Teotia, and Ostadabbas (2022a).

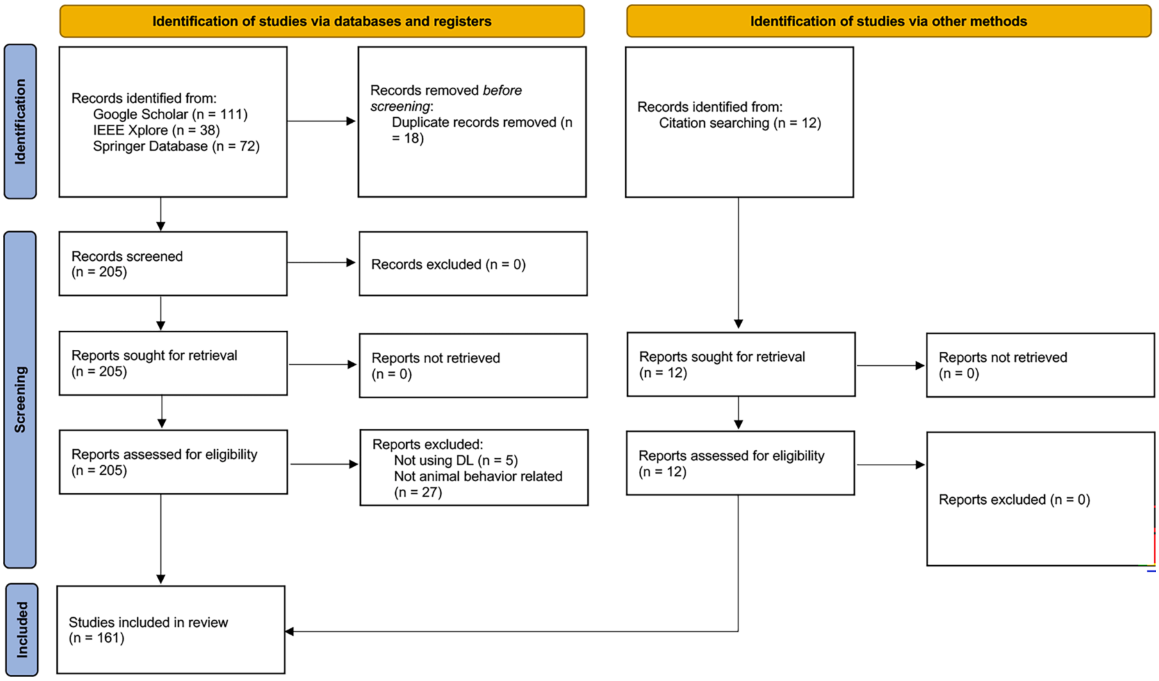

Fig. 1. PRIMA flowchart regarding our literature search.

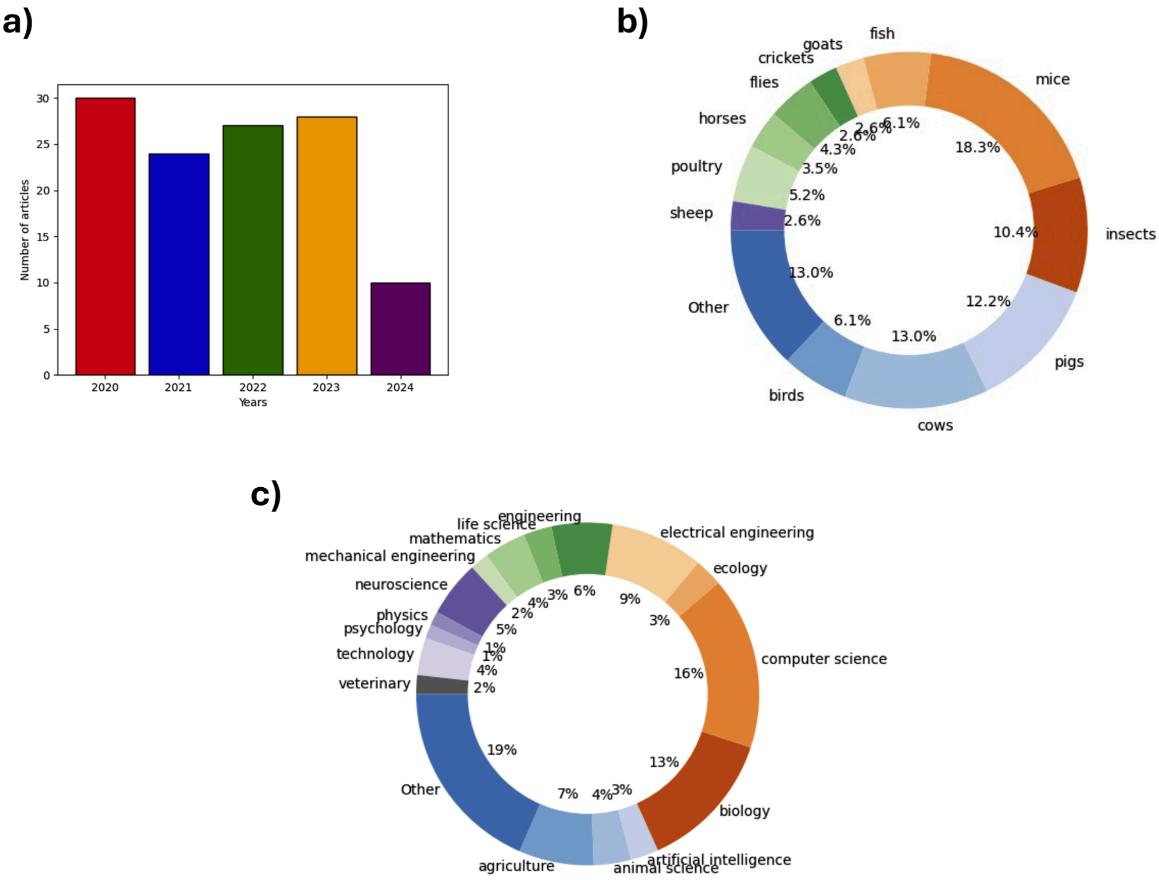

Fig. 2. (a) illustrates a histogram depicting the distribution of research articles per year, focusing exclusively on papers obtained and cataloged during the initial scavenging phase. (b) presents a pie chart detailing the variety of animals utilized in behavioral studies leveraging deep learning techniques. (c) displays another pie chart showcasing the diverse research fields of the authors.

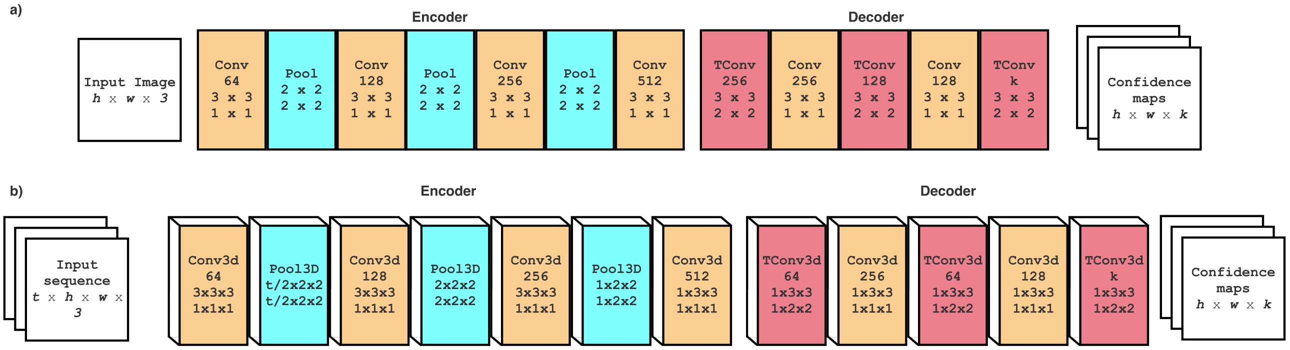

Fig. 3. (a) shows the architecture of LEAP Pereira et al. (2019); (b) the one exploited in T-LEAP Russello et al. (2022).

LEAP (Pereira et al., 2019) is a single-animal pose estimation model employing convolutional layers culminating in confidence maps that delineate the probability distribution for each distinct body keypoint. This architectural design, depicted in Fig. 3, is characterized by its simplicity, featuring three sets of convolutional layers. The initial two sets are terminated by *max pooling* to alleviate computational complexity. Subsequently, transposed convolution is applied to restore the original dimensions of the images, yet with a depth corresponding to the number of keypoints, thereby generating a confidence map for each. Despite its simplicity, the LEAP model encounters challenges in non-laboratory settings due to issues such as occlusion, prompting the introduction of T-LEAP (Russello, van der Tol, & Kootstra, 2022). T-LEAP preserves the architecture of LEAP but diverges in its use of 3D convolution instead of 2D convolution. The input to T-LEAP comprises four consecutive frames extracted from videos, enhancing the model's robustness. Notably, T-LEAP maintains a focus on single-animal pose estimation, as elucidated in Fig. 3. Subsequently, the author of LEAP introduced a refined version known as Social LEAP (SLEAP) (Pereira et al., 2022), designed to proficiently address the challenges associated with multi-animal pose estimation through the integration of both bottom-up (Papandreou et al., 2018) and top-down strategies (Nguyen & Kresovic, 2022). In the top-down strategy, SLEAP first identifies individuals and subsequently detects their respective body parts. Unlike LEAP, SLEAP seamlessly incorporates this approach without the need for an additional object detection architecture. On the other hand, the bottom-up strategy in SLEAP involves detecting individual body parts and subsequently grouping them into individuals based on their connectivity. A key advantage of this dual-strategy framework is its efficiency, requiring only a single pass through the neural network. The output of this strategy produces multi-part confidence maps and part affinity fields (PAFs) (Cao, Simon, Wei, & Sheikh, 2017), constituting vector fields that intricately represent spatial relationships between pairs of body parts. Additionally, SLEAP undergoes a structural enhancement by transitioning from LEAP's backbone to a more intricate U-Net architecture (Ronneberger, Fischer, & Brox, 2015), thereby significantly improving accuracy in the realm of multi-animal pose estimation scenarios.

Similarly, DeepLabCut (DLC) has evolved significantly over time. Initially designed as a single-animal pose estimation method, it utilizes a pretrained ResNet-50 (He, Zhang, Ren, & Sun, 2016) backbone with subsequent deconvolutional layers to generate confidence maps for keypoints. This approach, taking advantage of Imagenet pretrained weights, allowed DLC to effectively estimate skeletons with minimal data. The model's capabilities were later expanded to include 3D pose estimation through the use of multiple cameras (Nath et al., 2019). Each camera view was trained independently, and sophisticated camera calibration techniques were employed to derive 3D locations. A subsequent milestone in DLC's development involved addressing the challenges of multi-animal pose estimation (Lauer et al., 2022). This evolution introduced DLCRNet, a structural modification that replaced the ResNet backbone. DLCRNet employs a bottom-up multi-animal pose estimation approach, featuring a multi-fusion architecture and a multi-stage decoder. The decoder utilizes multiple stages of score maps and PAFs (Cao et al., 2017) to predict keypoints for each animal. Further innovation is exemplified by SuperAnimal (Ye et al., 2022), which introduced transformer layers into the model architecture.

While DLC and SLEAP currently stand as the predominant pose estimation methodologies in behavior analysis for animal behavior classification, it is imperative to acknowledge recent advancements in animal pose estimation architectures. Several notable methodologies have been introduced:

- **OptiFlex** (Liu et al., 2020b) is a video-based animal pose estimation method that, given a skip ratio s and a frame range f, assembles a sequence of $2f + 1$ images with indices ranging from $t - s \times f$ to $t + s \times f$. This sequence is input to a model based on residual blocks with intermediate supervision, generating predictions for each image and producing a sequence of heatmap tensors. These tensors are then fed into an OpticalFlow model, ultimately yielding the final heatmap prediction for index t. OptiFlex has demonstrated superior accuracy compared to DeepLabCut, LEAP, and DeepPoseKit (Graving et al., 2019).
- **SemiMultiPose** (Blau, Gebhardt, Bendesky, Paninski, & Wu, 2022) introduces a semi-supervised multi-animal pose estimation approach, building upon DeepGraphPose (Wu et al., 2020) and DirectPose (Tian, Chen, & Shen, 2019). Taking both labeled and unlabeled frames as input, the method processes them using a ResNet backbone, generating a compact representation fed into three branches: one for detecting keypoint heatmaps (B1), one for bounding box heatmaps (B2), and a third for keypoint detection (B3). SemiMultiPose aims to generate pseudo keypoint coordinates from B2 and B3 for the self-supervised branch, contributing to B1. The network has shown improved accuracy compared to SLEAP. However, the authors note that in cases of abundant labeled data, their method may not significantly outperform others, and for single-animal pose estimation with unlabeled frames from a sequential video, DeepGraphPose might outperform SemiMultiPose, benefiting from the consideration of spatial and temporal information.
- **Lightning pose** (Biderman et al., 2023) exploits spatiotemporal statistics of unlabeled videos in two ways. Firstly, it introduces unsupervised training objectives penalizing the network for predictions violating the smoothness of physical motion, multiple-view geometry, or departing from a low-dimensional subspace of plausible body configurations. Secondly, it proposes a novel network architecture predicting poses for a given frame using temporal context from surrounding unlabeled frames. The resulting pose estimation networks exhibit superior performance with fewer labels, generalize effectively to unseen videos, and provide smoother and more reliable pose trajectories for downstream analysis (e.g., neural decoding analyses) compared to previously mentioned approaches.

- Bhattacharya and Shahnawaz (2021) introduced a novel model for recognizing the pose of multiple animals from unlabeled data. The approach involves the removal of background information from each image and the application of an edge detection algorithm to the body of the animal. Subsequently, the motion of the edge pixels is tracked, and agglomerative clustering is performed to segment body parts. In a departure from previous methods, the end result is not specific keypoints but rather the segmentation of body parts. To achieve this, the authors utilized contrastive learning to discourage the grouping of distant body parts together.

After obtaining the skeletal representation of each animal in every frame, whether from videos or images, the subsequent step involves processing the data to discern specific behaviors. The trajectories derived from pose estimation can be effectively analyzed through statistical methods. Weber, Mulders, Kaiser, Tackenberg, and Rust (2022) utilized DeepLabCut predictions (Mathis et al., 2018) and ANOVA (Kaufmann & Schering, 2014) to conduct behavioral profiling of rodents, with a focus on studying stroke recovery. In a similar vein, Lee et al. (2021), employing DLC, investigated the behavior of non-tethered fruit flies. Their study involved predicting locomotion patterns and identifying the centroid of the animals' legs.

Machine learning for analyzing pose estimation trajectories becomes crucial when classifying postures and relating them to specific behaviors. One of the simplest approaches is to use a Nearest-Neighbor classifier. Saleh et al. (2023) tested this method to classify mouse behaviors such as crossing and rearing in an open-field experiment, achieving 97 % accuracy. Other machine learning approaches were employed by Fang, Zhang, Zheng, Huang, and Cuan (2021) and Nilsson et al. (2020). The former used a naive Bayesian classifier to identify eating, preening, resting, walking, standing, and running behaviors for poultry analysis, providing a disease warning system. The latter introduced the SimBa toolkit, importing DeepLabCut or DeepPoseKit projects to create classifiers using RandomForest Breiman (2001) and extracting features like velocities and total movements. Another application of Random Forest was employed by Higaki et al. (2024) for classifying on a scoring system basis dairy cows mobility. McKenzie-Smith, Wolf, Ayroles, and Shaevitz (2023) used trajectories obtained with SLEAP to identify stereotyped behaviors such as grooming, proboscis extension, and locomotion in *Drosophila melanogaster*, using resulting ethograms provided by MotionMapper (Berman, Choi, Bialek, & Shaevitz, 2014) to explore how flies' behavior varies across time of day and days, finding distinct circadian patterns in all stereotyped behaviors.

Other authors opted for recurrent and convolutional neural networks, with simple approaches such as using Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) and 1D convolutional neural networks to process trajectories for drawing behavioral conclusions. Examples include detecting lameness in horses (Alagele & Yildirim, 2022) and determining chemical interactions experienced by crickets (Fazzari, Carrara, Falchi, Stefanini, & Romano, 2024). More complex approaches include Wittek, Wittek, Keibel, and Güntürkün (2023)'s use of InceptionTime (Ismail Fawaz et al., 2020), an ensemble of deep convolutional neural network models, to classify seven distinct behaviors in birds. Some authors simplified the classification process by introducing a nonlinear clustering phase to improve the feature space, followed by classification using Multilayer Perceptrons (MLP), demonstrating advantages in classification (Schneider, Lee, & Mathis, 2023; Ye et al., 2022).

A recent emerging trend involves the utilization of unsupervised learning techniques in the analysis of animal behavior. Luxem et al. (2022) have innovatively proposed a methodology for processing trajectories derived from DeepLabCut by employing a Variational Auto-Encoder (VAE) (Kingma & Welling, 2013). Subsequently, they apply a Hidden Markov Model (HMM) (Rabiner & Juang, 1986) to the new representation of trajectories to discern underlying motifs. Following a comprehensive analysis of motif usage, the authors iteratively employ HMM, limiting the number of motifs to those surpassing a 1 % usage threshold in the previous analysis. The refined motifs were attributed to specific behavior exhibited by the mice, such as exploration, rearing, grooming, pausing, or walking. Notably, this methodological approach outperforms conventional techniques, such as Auto-Regressive HMM (AR-HMM) or MotionMapper (Berman et al., 2014), when applied directly to the motion sequences.

Motion trajectories extend their utility beyond predicting the behavior of individual animals; in multi-animal scenarios, they can also be applied to unravel the intricate web of social interactions among them. Segalin et al. (2021) introduced the Mouse Action Recognition System (MARS), a sophisticated automated pipeline tailored for pose estimation and behavior quantification in pairs of freely interacting mice. MARS adeptly discerns three specific social behaviors: close investigation, mounting, and attack. On a different note, Zhou et al. (2022) proposed the Cross-Skeleton Interaction Graph Aggregation Network (CS-IGANet), a groundbreaking framework designed to capture the diverse dynamics of freely interacting mice. CS-IGANet successfully identifies a spectrum of behaviors, including approaching, attacking, chasing, copulation, walking away from another mouse, sniffing, and many others.

Trajectories not only serve as a means to identify specific behaviors but are also instrumental in anomaly detection. For instance, Fujimori, Ishikawa, and Watanabe (2020) employed OneClassSVM (Boser, Guyon, & Vapnik, 1992) and IsolationForest (Liu, Ting, & Zhou, 2008) to detect outlier behaviors in domestic cats. Similarly, Gnanasekar et al. (2022) utilized pose estimation data to predict abnormal behavior in mice undergoing opioid withdrawal, employing pretrained convolutional neural networks for the classification of shaking behaviors.

For a comprehensive summary of the various methods discussed in this section, please refer to Table 3, which outlines each method along with its advantages, limitations, and potential applications.

5. Non pose estimation-based methods

In this section, we expound upon methodologies employed in the investigation of animal behaviors without recourse to pose estimation techniques. To enhance clarity and systematic presentation, we have delineated subsections corresponding to each methodology.

5.1. Sensor based approaches

Sensor-generated data, typically originating from accelerometers or gyroscopes, has been extensively explored in the literature, as comprehensively in Kleanthous et al. (2022b); Neethirajan (2020) surveys. These surveys delve into the application of classical machine learning methods in modern animal farming and the study of animal behavior. More recently, a shift towards leveraging deep learning approaches has been observed. Arablouei et al. (2023a) utilized a wearable collar tag equipped with an accelerometer to collect data from grazing beef cattle. They applied a Multi-Layer Perceptron to classify behaviors such as grazing, walking, ruminating, resting, and drinking. Similarly, Eerdekens et al. (2020) employed tri-axial accelerometers on horses, strategically positioned at the two front legs' lateral side. They proposed a Convolutional Neural Network to detect behaviors like standing, walking, trotting, cantering, rolling, pawing, and flank-watching based on the acquired data. Mekruksavanich et al. (2022), instead, segmented accelerometer data into 2-second windows and exploited a pre-trained ResNet model to perform sheep activity recognition. Dang et al. (2022) introduced the integration of multiple sensors, collecting environmental data (e.g., temperature, humidity) alongside cow behavior information obtained from accelerometers and gyroscopes. They preprocessed this information using a 1D-convolutional neural network and LSTM networks for classifying walking, feeding, lying, and standing. In a recent study, Pan, Chen, Zhong, Wang, and Zheng (2023) introduced four novel Convolutional Neural Network architectures tailored for Animal Action Recognition (AAR). These architectures, namely one-channel temporal

Table 3
Pose estimation techniques and associated processing methods applied to the keypoint location data obtained from the mentioned pose estimation model.

Method	Advantages	Limitations	Applications
LEAP (Pereira et al., 2019)	Simple and efficient model	Single-animal pose estimation, problems in non-laboratory settings due to occlusions	Single-animal keypoint estimation
T-LEAP (Russello et al., 2022)	Considers previous and subsequent frames for improving keypoint estimation	Single-animal pose estimation, requires images from a video to work	Single-animal keypoint estimation
SLEAP (Pereira et al., 2022)	Multi-animal pose estimation, both bottom-up and top-down approaches	No significant limitations	Single/Multi-animal keypoint estimation
DeepLabCut (Mathis et al., 2018)	Multi-animal pose estimation (DLCRNet), strong community, 3D location capabilities, various backbones (including Transformers, SuperAnimal)	Limited to bottom-up multi-animal pose estimation	Single/Multi-animal keypoint estimation
OptiFlex (Liu et al., 2020b)	Integrates optical flow to improve performance	Requires images from a video to create optical flow	Single-animal keypoint estimation
SemiMultiPose (Blau et al., 2022)	Uses semi-supervised learning	In cases of abundant labeled data, may not significantly outperform other single-animal pose estimation methods	Single-animal keypoint estimation
Lightning Pose (Biderman et al., 2023)	Uses spatiotemporal statistics to improve results, useful for unlabeled videos	Limited to videos	Single-animal keypoint estimation
Bhattacharya and Shahnawaz (2021)	Works with unlabeled data	Segments body parts, does not individuate specific keypoints	Body-part segmentation
SBeA (Han et al., 2024)	Requires fewer annotated samples for multi-animal pose estimation compared to DLC and SLEAP, enables 3D pose reconstruction	Requires a heavy data augmentation process	Single/Multi-animal keypoint estimation
ANORA	Robust for detecting unusual movements in animal behaviors	May struggle with highly variable behaviors	Behavioral profiling
Nearest-Neighbor Classifier	Simple, interpretable, works well with small datasets	Computationally expensive for large datasets, sensitive to noise	Animal behavior classification
Naive Bayesian Classifier	Fast, interpretable, works well with limited data	Assumes feature independence, limiting accuracy	Animal behavior classification (with probabilistic modeling)
Random Forest	Explainable	May not be optimal compared to deep learning approaches when the number of samples is huge	Feature extraction, classification
LSTM	Good for sequential data, captures temporal dependencies	Requires large datasets, computationally expensive	Animal behavior classification
1D Convolutional Neural Network	Efficient for time-series pose data	Less effective than LSTMs for long-range dependencies	Animal behavior classification
InceptionTime (Ismail Fawaz et al., 2020)	Proven to be better than LSTM and 1D CNN	Requires large training data and is more complex than LSTM and 1D CNN, since it uses ensembling and requires non-linear clustering and MLP training	Animal behavior classification
VAME (Luxem et al., 2022)	Researchers do not need to state how many behaviors are present a priori	Requires careful tuning and probabilistic interpretation	Unsupervised behavior discovery
MARS (Segalin et al., 2021)	Extracts precise movement features (speed, acceleration, etc.), useful for unsupervised analysis	Limited to predefined movement features	Quantitative analysis of animal motion
CS-IGANet (Zhou et al., 2022)	Captures social behaviors between two mice	Very complex architecture, requiring a lot of training data. Proven to work only on mice	Identification of social behaviors
OneClassSVM (Boser et al., 1992)	Good for outlier detection in pose-based behaviors	Sensitive to hyperparameters, may not generalize well	Detection of abnormal movements
IsolationForest (Liu et al., 2008)	Efficient anomaly detection	May not capture complex sequential dependencies	Identification of unusual movement patterns

(OCT), one-channel spatial (OCS), OCT and spatial (OCTS), and two-channel temporal and spatial (TCTS) networks, leverage data from 3D accelerometers and 3D gyroscopes. The core objective of their research was to scrupulously identify behaviors such as movement, drinking, eating, nursing, sleeping, and lying in lactating sows. More advanced techniques were employed by Otsuka et al. (2024) to classify wild animal behaviors using animal-borne accelerometers, leveraging Transformer architectures. Their attention mechanism allows for a more effective learning of temporal dependencies, enhancing classification performance.

In addition to accelerometer and gyroscope data, GNSS (Global Navigation Satellite System) data emerges as a valuable tool for understanding animal behavior. Arablouei, Wang, Bishop-Hurley, and Liu (2023b) explored this avenue by employing GNSS to extract pertinent information about cattle behavior, including metrics like distance from water points, median speed, and median estimated horizontal position error. Integrating this GNSS data with accelerometry information, the researchers pursued two distinct approaches. The first involved concatenating features from both sensor datasets into a comprehensive feature vector, subsequently fed into a MLP classifier. Alternatively, the second approach centered on fusing the posterior probabilities predicted by two separate MLP classifiers. These methodologies enabled the accurate detection of behaviors such as grazing, walking, resting, and drinking.

In conclusion, sensor-only data-based approaches are limited to uni- or multi-dimensional sequence analyses, depending on the type of sensor used. This makes such data comparable to pose estimation methods. In fact, the techniques discussed here share similarities with those used for processing pose estimation data, such as LSTMs, 1D CNNs, and Transformers. However, unlike pose estimation, which relies on capturing images or videos–a nearly cost-free process today due to the widespread availability of high-resolution smartphone cameras–sensor-based approaches introduce an additional expense related to sensor acquisition. Conducting an experiment requires purchasing multiple sensors, and if multiple animals need to be monitored simultaneously, a separate set of sensors must be obtained for each animal, further increasing costs.

The emergence of low-cost miniature sensors has helped mitigate these expenses (Jin et al., 2024), and researchers have explored methods to minimize the number of sensors needed by quantifying each sensor's contribution to accurate behavior identification. Addressing this concern, Li, Yang, Su, Li, and Wang (2023) introduced a novel method aimed at optimizing sensor selection. Their approach involves assessing the self-information brought by the jth sensor concerning the occurrence of a specific activity A_i and multiplying it by the universality of the same sensor j during instances of the same activity A_i. This innovative strategy has proven highly effective, leading to improved recognition rates and reduced computational time by eliminating redundant and noisy data, thereby lowering overall costs.

5.2. Bioacoustics

While bioacoustics offers a captivating glimpse into animal behavior (Stowell, 2022) and their ecosystem (Oestreich, Oliver, Chapman, Go, & McKenna, 2024), given the integral role of sound in animal activities such as communication, mating, navigation, and territorial defense (Chalmers, Fergus, Wich, & Longmore, 2021), the current landscape of published articles predominantly emphasizes animal identification (Bravo Sanchez, Hossain, English, & Moore, 2021; Varma, Bateshwar, Rathi, & Singh, 2021; Xu, Zhang, Yao, Xue, & Wei, 2020) and sound event detection (Moummad, Serizel, & Farrugia, 2023; Nolasco et al., 2023). Notably, the existing literature reveals a scarcity of research endeavors combining acoustics and deep learning for the identification of animal behaviors. Wang, Wu, Cui, Xuan, and Su (2021a) stand out as pioneers in this domain, as they endeavored to classify sheep behaviors, including chewing, biting, chewing-biting, and ruminating sounds. This was accomplished using a recording device positioned proximal to the animal's face, with a placebo class designated as noise. The acquired wavelet data were leveraged for classification tasks through both a feed-forward neural network and a recurrent neural network. Additionally, the information was further processed by transforming it into a log-scaled Mel-spectrogram, serving as input for a convolutional neural network. The findings underscore that while the recurrent neural network exhibited superior performance, the convolutional neural network outperformed the feed-forward approach, attributing its success to the enhanced signal representation offered by the Mel-spectrogram.

CNNs have gained increasing attention in animal tasks, particularly due to the benefits of transfer learning, which enhances their performance on new datasets. Notable examples include the work of Manriquez P, Kotz, Ravignani, and De Boer (2024), who used CNNs to classify mammals vocalizations, and Schall, Kaya, Debusschere, Devos, and Parcerisas (2024), who applied them to detect baleen whales. Therefore, we believe that transfer learning could also improve behavior identification through sound, making it a valuable and effective approach to consider.

5.3. Object detection

In conjunction with pose estimation techniques, object detection stands out as a widely employed deep learning methodology for analyzing animal behavior. Its prevalence may be attributed to its established utility in animal recognition and detection (Banerjee, Khan, & Sharma, 2023; Chen, Zhu, & Norton, 2021; Teixeira et al., 2023), prompting researchers to redirect their focus toward studying animal welfare and activity.

Among the leading architectures for animal behavior identification, Faster R-CNN (Ren, He, Girshick, & Sun, 2015) and particularly YOLO (Redmon, Divvala, Girshick, & Farhadi, 2016) are frequently employed. Faster R-CNN follows a two-stage approach, first generating region proposals and then refining predictions, which provides high detection accuracy but comes at the cost of increased computational complexity, making it less suitable for real-time applications. In contrast, YOLO adopts a single-stage approach, directly predicting object classes and bounding boxes in one pass through the network. This design significantly enhances processing speed, making YOLO more favorable for real-time animal behavior analysis, albeit sometimes at the expense of detection precision. Alternative architectures have also been proposed. For instance, Samsudin, Harizan, Ibrahim, Karim, and Ibrahim (2022) utilized SSD MobileNetv2 (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2018), a lightweight and efficient model, to detect abnormal and normal zebrafish larvae behaviors for examining the effects of neurotoxins. However, SSD MobileNetv2 typically trades off some accuracy for efficiency. To address spatiotemporal dependencies in behavioral analysis, McIntosh, Marques, Albu, Rountree, and De Leo (2020) introduced TempNet, which incorporates an encoder bridge and residual blocks with a two-stage spatial-temporal encoder. This architecture enhances the detection of dynamic behaviors, such as startle responses in fish, by capturing motion patterns more effectively than frame-based detectors like YOLO and Faster R-CNN.

Object detection serves a dual role, encompassing instantaneous behavior detection through image or video frame analysis, as well as the quantification and tracking of specific behaviors. The accurate analysis of single frames, counting, and frame-by-frame examination enable researchers to quantify both the duration and frequency of distinct actions. For instance, the application of YOLO in the study by Alameer et al. (2022) facilitated the quantification of contact frequency among pigs, allowing the identification of peculiar behaviors such as rear snorting and tail-biting. In the context of cows and pigs, a crucial aspect involves quantifying movement and aggressive behavior (Alameer et al., 2022; Odo, Muns, Boyle, & Kyriazakis, 2023). Furthermore, efforts to discern rank relationships based on fighting behavior in animals like cows are of crucial importance (Uchino & Ohwada, 2021). Importantly, for tasks demanding prolonged animal identification, tracking is conventionally executed using DeepSort (Evangelista, Concepcion, Palconit, Bandala, & Dadios, 2022; Wojke, Bewley, & Paulus, 2017).

Efficient instant detection can be accomplished by conducting a single analysis on the animal and directly classifying its behavior through a single image. In this context, deep learning object detection models prove instrumental in directly identifying behaviors such as positional activities (e.g., mating, standing, feeding, spreading, fighting, drinking) for the comprehensive analysis of animal health and stress behaviors (Manoharan, 2020; Riekert, Klein, Adrion, Hoffmann, & Gallmann, 2020; Wang, Wang, Li, & Ren, 2020). These models also find application in disease identification, such as the detection of wryneck (Elbarrany, Mohialdin, & Atia, 2023), and in studying behavioral adaptations to new environments (Li et al., 2019a). Furthermore, object detection models can be extended to operate with thermal and infrared images, which are getting more popular (Korelidou, Simitzis, Massouras, & Gelasakis, 2024). For example, Xudong, Xi, Ningning, and Gang (2020) utilized thermal images for the automatic recognition of dairy cow mastitis, introducing the EFMYOLOv3 model. Similarly, Lei et al. (2022) employed infrared images to discern feeding, resting, moving, and socializing behaviors in slow animals.

Beyond these applications, notable approaches utilizing object detection include Fuentes, Yoon, Park, and Park (2020), who integrated YOLO and Optical Flow to detect actions in cows. Additionally, some researchers employ object detection solely for localizing the animal within the image or video. They subsequently crop that region and use it in other models, leveraging 2D pretrained networks or introducing 3D convolutional neural networks for video analysis (Feighelstein et al., 2023; Thanh & Netramai, 2022).

5.4. Others

Several research endeavors have employed unique deep learning methodologies, distinct from those discussed in the preceding section. Due to the relative scarcity of deep learning strategies for evaluating animal behavior in the existing literature using the aforementioned approaches, we endeavored to compile a comprehensive assortment of ideas. To achieve this, we have identified and categorized five distinct approaches:

- **Convolutional lassification on raw data.** Alameer, Kyriazakis, Dalton, Miller, and Bacardit (2020) employed a GoogLeNet-like architecture (Szegedy et al., 2015) to discern between feeding and non-nutritive visits to a manger in pig recordings. In a similar vein, Ayadi et al. (2020) utilized VGG19 (Simonyan & Zisserman, 2014) to determine whether cows were ruminating or not. This network architecture was also applied to identify various behaviors in mice, such as grooming, licking the abdomen, squatting, resting, circling, wandering, climbing, and searching (Wang et al., 2021b). Similarly leveraging pre-trained networks, Andresen et al. (2020) developed a fully automated system for surveilling post-surgical and post-anesthetic effects in mice facial expressions, employing InceptionV3 (Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2016) for pain identification. Notably, Bohnslav et al. (2021) introduced DeepEthogram, a software tested for predicting mice and flies behaviors. The approach involves using a sequence of 11 frames, where the last frame is the target for prediction. Optical flow frames are generated using MotionNet (Zhu, Lan, Newsam, & Hauptmann, 2019). These frames, along with the target frame, are fed into a feature extractor (ResNet architectures (Hara, Kataoka, & Satoh, 2018; He et al., 2016)) to extract both flow and spatial features. Subsequently, the features are concatenated, and Temporal Guassian Mixture (TGM) model (Piergiovanni & Ryoo, 2019) is applied for classification. Han, Zhu, Liu, Zhang, and Xie (2020) employed a simpler approach, superimposing the frame to be predicted with computationally generated optical flow from the subsequent frame. The resulting image is then classified using a convolutional neural network to categorize behaviors in fish shoals, including normal state, group stimulated, individual disturbed, feeding, anoxic, and starvation state.
- **Segmentation.** Xiao et al. (2023) employed Mask R-CNN (He, Gkioxari, Dollár, & Girshick, 2017) to segment birds within a 3D space, facilitating the analysis of their interactions based on distinct social actions: approach, stay, leave, and sing to. In contrast, other researchers have devised innovative pipelines to investigate animal behavior. EthoFlow (Bernardes et al., 2021) is a software grounded in segmentation, enabling the tracking and behavioral analysis of organisms (validated on bee datasets). On a different note, SIPEC (Marks et al., 2022) constitutes a pipeline leveraging an Xception network (Chollet, 2017) to extract features from frames. These features are subsequently processed over time using a Temporal Convolutional Network (TCN) (Lea, Vidal, Reiter, & Hager, 2016) to classify the animal's behavior in each frame. While SIPEC abstains from segmentation in the case of single animal classification, it seamlessly incorporates segmentation for multi-animal behavior classification. Segmentation, hence, is limited to provided extract valuable features to be processed afterwards.
- **Self-supervised learning.** Jia et al. (2022) proposed an innovative self-supervised learning approach known as Selfee, designed for extracting comprehensive and discriminative features directly from raw video recordings of animal behaviors. Selfee utilizes a pair of Siamese convolutional neural networks (Koch, Zemel, & Salakhutdinov, 2015), trained explicitly to generate discriminative representations for live frames. The authors highlighted that their method effectively captures global behavioral characteristics, making it resilient to occlusions. However, compared to keypoints obtained through pose estimation, Selfee's learned features are less interpretable. Despite this limitation, the model is computationally efficient and can be trained on low-cost GPUs, making it more accessible to biology labs with limited computational resources. Nonetheless, the authors acknowledged that relying solely on three raw frames per live input may limit the model's ability to fully capture the nuances of animal motion.
- **Explainability.** To the best of our knowledge, Choi, Pyenson, Liebig, and Pavlic (2022) stands as the sole contributor employing Explainable Artificial Intelligence (XAI). In their study, they harnessed Grad-CAM (Selvaraju et al., 2016) to delve into the decision-making process of a neural network designed to distinguish between unstable and stable ant swarms. The investigation aimed to ascertain the network's capacity to comprehend intricate behaviors such as dueling and dominance biting, shedding light on the explainability of its predictions.
- **Behavior identification in clips.** Li, Zhang, Li, and Chen (2020) undertook the task of categorizing significant pig behaviors, such as feeding, lying, motoring, scratching, and mounting. Their approach involved the development of Pig's Multi-Behavior recognition (PMB-SCN), a sophisticated architecture built upon the SlowFast framework (Feichtenhofer, Fan, Malik, & He, 2019) and leveraging spatio-temporal convolution. PMB-SCN comprised two distinct SlowFast pathways with varying temporal speeds. The *slow pathway* utilized a larger temporal stride when processing input frames (e.g., 8, considering a clip with a length of 64 frames), while the *fast pathway* employed a smaller temporal stride (e.g., 2). The features extracted by these pathways were interconnected through lateral connections (Lin et al., 2017), enhancing the model's ability to capture complex spatio-temporal patterns. The final phase of the methodology involved classification, where the fused features were utilized to discern and categorize various pig behaviors effectively. Recently, also multimodal action identification on multiple animals was performed leveraging information from the video, image and label data, using a model called MSQNet (Mondal, Nag, Prada, Zhu, & Dutta, 2023). Multimodal models have demonstrated excellent capabilities in detecting a wide range of actions in video clips. However, their increased network complexity can hinder real-time analysis due to slower processing times. To address this limitation in MSQNet, a new version called Mamba-MSQNet was introduced by Fazzari, Romano, Falchi, and Stefanini (2025). This version replaces the Transformer components with Mamba layers (Gu & Dao, 2023), significantly reducing both the number of parameters and FLOPs, thus improving computational efficiency, without reducing performance.

Concluding, all non-pose estimation-based methods described in this section are summarized in Table 4, along with their advantages, limitations and how they are applied.

6. Publicly available datasets

This section meticulously enumerates publicly accessible datasets featured or referenced in the articles identified through our systematic search. Table 5 presents details on each dataset, including the article of introduction, authorship, targeted species, data type (e.g., images, videos, audio signals, sensor data), the specific tasks for which they were utilized and the content of the datasets. Noteworthy is the incorporation of references indicating the dual usage of datasets–initially introduced for a specific task and subsequently repurposed, signified by citations in the *Application* column. Regrettably, several articles utilized private datasets, although some authors may offer dataset-sharing options. We recommend consulting the corresponding articles to explore potential data access avenues.

A number of key considerations can be gleaned from Table 5, as follows:

- In the context of datasets tailored for pose estimation, a salient observation is the standardization of skeletal structures when deployed across diverse animal species (Cao et al., 2019; Ng et al., 2022; Yu et al., 2021). This standardization facilitates the training of a singular network, avoiding the need for species-specific networks.

Table 4
Non-pose estimation techniques methods applied to animal behavior analysis.

Method	Advantages	Limitations	Applications
MLP	Simple model	Do not consider temporal and spatial information, too simplistic in most applications	Animal behavior classification on sensor data
LSTM	Good for sequential data, captures temporal dependencies	Less effective for very long-range dependencies compared to Transformers	Animal behavior classification on sensor data, bioacoustics
1D CNN	More efficient to LSTM for time-series	Less effective than LSTMs for long-range dependencies	Animal behavior classification on sensor data
TCTS (Pan et al., 2023)	Convolutional neural network using 3D sensors data	Limited adaptability to highly variable time-series patterns	Animal behavior classification on sensor data
Transformer	Can handle long-range dependencies, either convolutional or not	Requires large datasets and high computational resources	Animal behavior classification on sensor data, images, video
CNN	Works with images, spectrograms, and pre-trained architectures, making training easy	Captures only spatial dependencies, lacks temporal modeling	Animal behavior classification on bioacoustics data, images
Faster R-CNN (Ren et al., 2015)	Good performance for object detection	Two-stage detection makes it computationally expensive	Animal localization
SDD MobileNetv2 (Sandler et al., 2018)	Lightweight object detection model	Reduced performance compared to newer versions of YOLO	Animal localization
YOLO (Redmon et al., 2016)	Fast real-time object detection thanks to single-stage approach, extended to video by integrating Optical Flow	Single-stage approach may reduce accuracy	Animal localization
TempNet (McIntosh et al., 2020)	Better for dynamic behaviors than YOLO and Faster R-CNN	No important limitations	Animal localization
EFMYOLOv3 (Xudong et al., 2020)	Works with Thermal images	No important limitations	Animal localization using thermal images
Mask R-CNN (He et al., 2017)	High accuracy in object segmentation	Segmentation cannot provide behavioral information, but can be used as feature	Segmentation
EthoFlow (Bernardes et al., 2021)	Specialized for behavioral segmentation	Segmentation cannot provide behavioral information, but can be used as feature	Segmentation
SIPEC (Marks et al., 2022)	Multi-animal behavioral segmentation	Segmentation cannot provide behavioral information, but can be used as feature	Segmentation
Selfee (Jia et al., 2022)	Self-supervised learning for feature extraction	Additional layer for classifying features is needed	Self-supervised learning (feature extraction)
Grad-CAM (Choi et al., 2022)	Explainable	Gives information about what the deep learning model consider for its classification, however the researcher has to understand why those part are meaningful	Understanding what the model distinguish as stable as non-stable swarms is related to some kind of collective behavior
PMB-SCN (Li et al., 2020)	Spatio-temporal convolution based on SlowFast framework	Model proposed specifically for pigs	Animal behavior classification in video clips
MSQNet (Mondal et al., 2023)	Multimodality helps in improving results, tested on Animal Kingdom capable of discerning 140 different actions spanning 850 species	Transformer architecture makes it complex to train, also multimodality add an increase in complexity	Multi-animal behavior recognition in video clips
Mamba-MSQNet (Fazzari et al., 2025)	Mamba architecture improves efficiency over MSQNet, Multimodality helps in improving results, tested on Animal Kingdom capable of discerning 140 different actions spanning 850 species	Requires a GPU to use Mamba efficiently	Multi-animal behavior recognition in video clips

Conversely, datasets exclusive to individual animal species exhibit intricate skeletal configurations tailored to the anatomical nuances of that species. For instance, precision in detecting the distal and proximal ends of crickets' antennae may be achievable (Fazzari, Carrara, Falchi, Stefanini, & Romano, 2023), but such granularity may not translate to not insect species like horses.

- The inclination of animals like goats and birds to move in open fields compels researchers to rely on sensor data or alternative methods, such as audio recordings, for event and action detection. This approach is significantly more manageable than tracking the animals with cameras. However, the utilization of sensors is constrained by the availability and affordability of these devices, directly impacting the number of individuals involved and the quantity of sequences that can be compiled for the dataset.
- For identifying static positions, such as whether an animal is lying down or standing, still images suffice. However, capturing and analyzing videos, or more precisely, short video clips, is crucial for recognizing dynamic actions and behaviors. These clips are intentionally brief to focus solely on the relevant action event, ensuring accurate classification using deep learning techniques. This approach effectively eliminates extraneous or unrelated behaviors that may interfere with the identification of the specific behavioral instance. This is the rationale behind Yang et al. (2022)'s decision to utilize 15 frames for each video clip.
- Unfortunately, publicly available collective and social behavior prediction and analysis datasets are currently limited to mice and fish, even though we discussed a study in this survey that utilized explainable artificial intelligence to analyze ant behavior. This innovative research methodology relies heavily on video data, presenting a computational challenge that demands substantial processing efforts. The intricate nature of this approach necessitates a considerable investment of time for meticulous frame and event labeling, thereby slightly diminishing its overall research appeal and popularity.
- An essential consideration pertains to the primary focus of many databases, which primarily aim at identification, detection, pose estimation, and tracking. Despite this orientation, it is crucial to acknowledge that several datasets have been instrumental in behavioral analysis, even if not explicitly designed for such purposes. Researchers are strongly encouraged not to overlook animal datasets

Table 5
Datasets Information useful for animal behavior analysis.

Species	Introduced by	Type	Applications	Dimensions
Baboons	Duporge et al. (2024)	Videos	Action recognition	23,409 video clips
Birds	Akçay et al. (2020)	Images	Detection	3436 images
Birds	Morfi, Bas, Pamuła, Glotin, and Stowell (2019)	Audio	Recognition (Bravo Sanchez et al., 2021)	687 recording, 87 classes
Birds	Knight and Bayne (2019)	Audio	Event detection (Lostanlen et al., 2019)	64 recording, 5 classes
Birds	Shamoun-Baranes, Burant, van Loon, Bouten, and Camphuysen (2017)	Sensor data	Movement prediction (Wijeyakulasuriya et al., 2020)	19 sequences
Cattle	Arablouei et al. (2023b)	Sensor data	Behavior classification	11,962 labeled datapoints (arm20c), 10,879 labeled datapoints (arm20e)
Dogs	Barnard et al. (2016)	Images	Pose Estimation	22,479 images
Goats	Kamminga et al. (2018)	Sensor data	Action recognition (also Bocaj, Uzunidis, Kasnesis, & Patrikakis, 2020)	177.8 hours of sequence data, 5 individuals
Goats	Kleanthous, Hussain, Khan, Sneddon, and Liatsis (2022a)	Sensor data	Action recognition (also Bocaj et al., 2020; Mekruksavanich et al., 2022)	2 sequences
Horses	Kamminga, Meratnia, and Havinga (2019)	Videos	Pose estimation, Action recognition (Bocaj et al., 2020)	8144 frames
Horses	Mathis et al. (2021)	Images	Pose Estimation	608,550 images
Fish	Mathis et al. (2018)	Videos	Pose estimation	100 frames
Fish	McIntosh et al. (2020)	Videos	Behavior classification	892 clips
Fish	Papaspyros et al. (2023)	Videos	Collective Behavior Prediction	three 16-hour trajectory datasets
Fish	Rahman, Song, Leung, Lee, and Lee (2014)	Videos	Action recognition (Gore, Kakodkar, Del Rosario Hernández, Edmister, & Creton, 2023)	95 clips
Fish	Tucker Edmister et al. (2022)	Videos	Tracking, Chemical response analysis (Gore et al., 2023)	3 hours for each of the 384 individuals
Insects	Bjerge et al. (2023)	Images	Detection	29,960 images
Insects	Fazzari et al. (2024)	Videos	Tracking, Chemical response analysis	5 minutes clip for 69 individuals
Insects	Modlmeier et al. (2019)	Images	Movement prediction (Wijeyakulasuriya et al., 2020)	14,400 frames (4 hours)
Insects	Pereira et al. (2022)	Videos	Pose estimation	30 videos, 2000 labels
Insects	Pham (2022)	Sequences	Locomotion classification	258 traces
Insects	Ullah et al. (2022)	Images	Recognition	1686 images
Jellyfish	Martin-Abadal, Ruiz-Frau, Hinz, and Gonzalez-Cid (2020)	Images	Detection	842 images
Mice	Burgos-Artizzu, Dollár, Lin, Anderson, and Perona (2012)	Videos	Pose estimation, Social behavior analysis (Jiang et al., 2021)	237 videos and over 8M frames
Mice	Jiang et al. (2021)	Videos	Social behavior analysis	12*3 annotated videos, 216,000*3 frames in total
Mice	Jiang et al. (2022b)	Videos	Detection, Tracking	10 videos, each video lasts 3 min, 4000 frames annotated
Mice	Mathis et al. (2018)	Videos	Pose estimation	161 frames
Mice	Pereira et al. (2022)	Videos	Pose estimation	1000 frames, 2950 instances & 1474 frames, 2948 instances
Mice	Segalin et al. (2021)	Videos	Pose estimation, Behavior classification	3.3M labels for pose annotation, 14 hours of behavior annotation
Multiple, 5	Cao et al. (2019)	Images	Pose Estimation	4666 images
Multiple, 11	Liu et al. (2023)	Images/Video	Pose estimation, Instance segmentation, Action recognition, Object Detection	35K images and 10K videos
Multiple, 24	Lu et al. (2023)	Images/Video	Recognition, Action recognition	2256 videos
Multiple, 30	Yang et al. (2022)	Images/Videos	Pose Estimation	2.4K video clips with 15 frames for each video, resulting in 36K frames
Multiple, 54	Yu et al. (2021)	Images	Pose Estimation	10,015 images
Multiple, 173	Chen et al. (2023)	Videos	Behavior recognition	18,346 videos
Multiple, 850	Ng et al. (2022)	Images/Videos	Pose Estimation, Action recognition	50 hours of annotated videos, 30K video sequences for AAR, 33K frames for APE
Pigs	Riekert et al. (2020)	Images	Position classification	7277 images
Tigers	Li, Li, Tang, Qian, and Lin (2019b)	Images	Pose estimation	8076 images
Monkeys	Labuguen et al. (2021)	Images	Pose estimation	13,083 images

merely because they do not pertain to specific behaviors. Valuable insights can be gleaned from these datasets, and their broader applicability should be explored beyond their initially intended scope.

7. Discussion about research questions

In concluding our extensive survey, we undertake the task of responding to the research questions outlined in Section 2.3, drawing upon the insights gleaned from the studies expounded upon earlier. These responses aim to offer readers practical guidance in navigating the dynamic trends discerned from the comprehensive examination of the state-of-the-art literature. Their purpose is to serve as a compass for readers, facilitating a deeper comprehension of emerging patterns and fostering the implementation of advancements in the field of animal behavior.

RQ1 (Which animal species are more considered and why?): In the context of animal behavior analysis employing deep learning, farm animals take center stage, as illustrated in Fig. 2. Pigs and cows emerge as prominent subjects, while mice claim a noteworthy position owing to their significance in neuroscience research (Bryda, 2013). Despite chickens being the most widely farmed animals globally, surpassing even cows, sheep, and goats (Robinson et al., 2014), their consideration in behavior analysis appears relatively subdued. This discrepancy may be

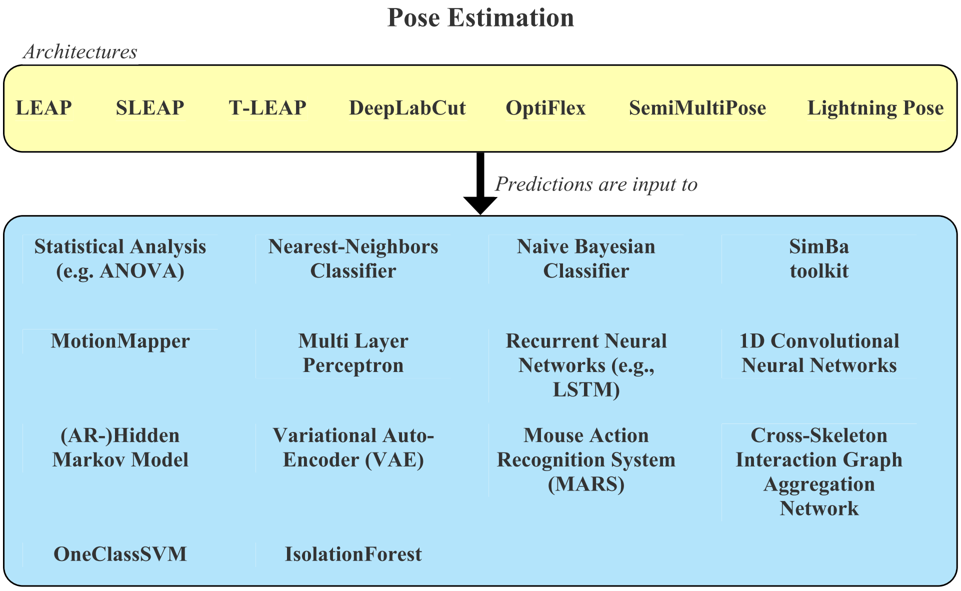

Fig. 4. Comprehensive schema illustrating the pose estimation architectures covered in this survey, accompanied by detailed methodologies for accurate classification of predictions into distinct behavioral classes.

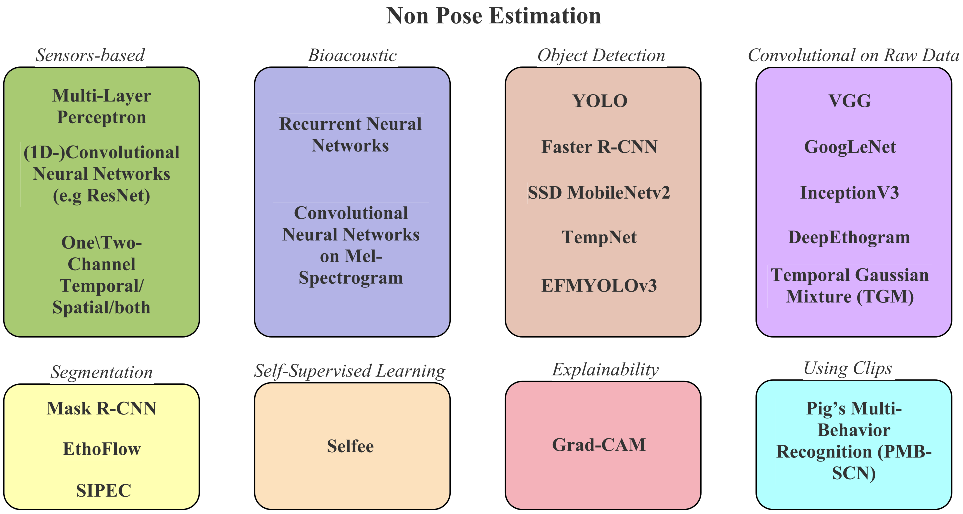

Fig. 5. Comprehensive schema illustrating the non-pose estimation architectures covered in this survey. To improve readability we divided them into blocks using the same structure employed in the survey.

attributed to the methods of chicken rearing, specifically the use of battery cages, which restrict movement and complicate behavioral tracking due to occlusion. Additionally, the large flock sizes typical of broiler production may limit the applicability of individual behavioral data, reducing its relevance for precise behavioral analysis. They are predominantly analyzed for welfare-related assessments (Joo, Duan, Weimer, & Teli, 2022; Mohialdin, Elbarrany, & Atia, 2023).

Goats and sheep, less scrutinized in behavioral studies, likely owe by their less financial importance compared to other animals and their lower visibility to their outdoor grazing habits, which hinder the feasibility of employing image processing techniques. Unlike pigs, which are often studied within confined spaces, the vast open areas in which goats and sheep are typically raised limit the practicality of utilizing image processing, even with occasional drone deployment (Al-Thani, Albuainain, Alnaimi, & Zorba, 2020). Birds face a similar challenge, requiring continuous tracking or data collection from sensors for comprehensive analysis (Bergen et al., 2022).

For aquatic creatures like fishes, a distinct hurdle arises from the difficulty in training neural networks on underwater images. These images often suffer from poor quality due to distortion and color/contrast loss in water, necessitating an image enhancement phase for meaningful analysis (Saleh, Sheaves, Jerry, & Azghadi, 2022).

A notable observation is the limited consideration given to domestic animals in this research context (Chambers et al., 2021; Choi, Chae, Lee, Park, & Chung, 2021; Kasnesis et al., 2022; Lecomte, Audet, Harnie, & Frigon, 2021). This may stem from the scarcity of veterinary professionals engaged in this evolving field or ethical concerns surrounding the study of domestic animals.

RQ2 (What deep learning methods have been used in the literature for animal behavior analysis?): Throughout this survey, we delineate two distinct methodologies employed for the analysis of animal behaviors: pose estimation and non-pose estimation methods. Pose estimation-based approaches hinge on the analysis of trajectories traced by keypoints, with subsequent utilization of various machine and deep learning techniques to scrutinize behavior. Predominantly, recurrent neural networks and 1D convolutional neural networks are the favored deep learning methods, though recent applications have also embraced variational auto-encoders for unsupervised motif identification (Kingma & Welling, 2013; Luxem et al., 2022) and convolutional graph networks for interaction analysis (Zhou et al., 2022). However, a prevalent trend emerges wherein data is often processed through statistical analysis or traditional machine learning methods (Fang et al., 2021; McKenzie-Smith et al., 2023; Nilsson et al., 2020; Saleh et al., 2023). This inclination may stem from the fact that many researchers are not inherently engaged in artificial intelligence or data science, as mentioned earlier (see Section 3.2), or it could be influenced by the volume of available data, given the heightened data requirements of deep learning (Özdaş, Uysal, & Hardalaç, 2023). Despite the migration of classification tasks and behavior discovery to deep learning, certain aspects, such as behavior outlier detection, persist in employing classical machine learning approaches like OneClassSVM (Boser et al., 1992) and IsolationForest (Liu et al., 2008). On the other hand, non-pose estimation encompasses diverse applications, categorized based on the type of data: sensors, audio and video, and image data. Sensor data commonly undergoes processing through (1D)-convolutional or recurrent neural networks (Dang et al., 2022; Eerdekens et al., 2020; Mekruksavanich et al., 2022; Pan et al., 2023), and in some cases, multi-layer perceptron classifiers (Arablouei et al., 2023a), particularly when sensor data is transformed into features like velocity, angles, humidity, location, among others (Arablouei et al., 2023b). In the domain of bioacoustics, recurrent neural networks or pre-trained convolutional neural networks on spectrogram images are frequently applied (Wang et al., 2021a). For images and videos, processing methods vary according to the task at hand. They may be employed for object detection, where identification extends beyond the animal to encompass specific behaviors, typically achieved through frameworks such as YOLO (Redmon et al., 2016) and Faster R-CNN (Ren et al., 2015). Alternatively, pre-trained neural networks or segmentation techniques, with subsequent analysis of the segmentation mask, are utilized (Xiao et al., 2023). Figs. 4 and 5 encapsulate and illustrate the summarized pose and non-pose estimation methods for behavior analysis expounded in this survey.

RQ3 (What are the deep learning strategies that are suitable and could enhance this task, but are not yet exploited?):

A key observation from the methodologies surveyed in this work is that many models used in animal behavior analysis are relatively outdated. Despite this, they remain widely adopted due to their extensive validation across various research domains, making them reliable and well-established approaches. Researchers who are not deep learning experts may prefer these models because they are easier to implement, thanks to the abundance of available literature and open-source code. A notable example is the persistent use of older YOLO models, particularly YOLOv3, despite the availability of more advanced versions. This highlights the need for a systematic update of models across different methods to incorporate state-of-the-art advancements. In contrast, action recognition models tend to adopt highly complex, cutting-edge architectures. This trend may be driven by a stronger emphasis on achieving state-of-the-art computer vision performance rather than addressing ethological concerns, as evidenced by the publication venues where these models are typically introduced.

Another crucial avenue for advancing deep learning in animal behavior analysis is the integration of Reinforcement Learning (RL). While RL has been previously employed in animal studies, its applications have primarily focused on emulating animal movement for robotic implementations (Peng et al., 2020). However, RL could be leveraged to create high-fidelity digital twins (Liu, Xu, Liu, & Wang, 2022), virtual representations of animals that mimic their behavior in response to various stimuli. These digital twin models offer a powerful tool for studying animal behavior in a controlled and scalable manner, eliminating the need for direct physical interaction. Developing such high-fidelity models presents challenges, but even approximate simulations that allow virtual animals to interact with stimuli, other animals, or objects could provide unprecedented insights into behavioral dynamics (Mori et al., 2022; Romano & Stefanini, 2021; Yamaguchi et al., 2018). This approach has already been explored in human studies, where simulated communities are used to analyze social behavior (Park et al., 2023b). Applying a similar strategy to animals could reveal subtle learned behaviors that emerge through unsupervised interactions. Additionally, digital twins could facilitate the development of AI agents capable of interacting with animals, learning how to influence their behavior through reinforcement-based strategies. Directly testing such interactions in real-world settings may be impractical or ethically concerning, as unintended harm could occur. However, a two-phase approach–first training an agent in a virtual environment and subsequently applying domain adaptation to real-world scenarios–could yield transformative results in animal behavior research. Ultimately, the fusion of deep learning and reinforcement learning holds the potential to create dynamic, interactive simulations that significantly enhance our understanding of animal behavior across diverse contexts.

Beyond these innovative applications, RL can also contribute to improving the general performance of deep learning models. RL has been successfully employed for optimizing neural network architectures (Jaafra, Laurent, Deruyver, & Naceur, 2019), tuning hyperparameters (Rijsdijk, Wu, Perin, & Picek, 2021), and dynamically enhancing data augmentation strategies to improve model accuracy (Liu et al., 2020a). Additionally, since some behavioral studies incorporate sensor data, RL can assist in intelligently selecting and integrating sensor inputs, reducing costs, improving efficiency, and eliminating redundancies (Tilak, Mukhopadhyay, Tuceryan, & Raje, 2010).

8. Conclusions

This survey examined the manifold benefits associated with the application of deep learning methodologies in the identification of animal behavior. We reviewed and analyzed a broad spectrum of methodologies, spanning sequence processing techniques, classification, anomaly detection, object recognition, and many others. Moreover, we curated a comprehensive table of publicly available datasets relevant to animal behavior, thereby augmenting the practical utility of deep learning applications.

Addressing our research questions, we found that farm animals (especially pigs, cows, and mice) are the most studied, due to their relevance in agriculture and neuroscience. Pose estimation and recurrent or convolutional neural networks dominate current methods, though classical approaches persist due to accessibility and data limitations. Emerging strategies like reinforcement learning and digital twins remain underexplored yet hold great promise for advancing the field. Our discourse on the subject and prospective considerations has pinpointed extant challenges within the literature, proffering a roadmap for potential research trajectories conducive to the advancement of the field. In essence, this survey serves as an invaluable compendium for researchers spanning diverse domains, with particular relevance to ethologists and neuroscientists. We believe that this survey will help inform forthcoming research initiatives and contribute to advancements in the intricate domain of animal behavior studies using deep learning.

CRediT authorship contribution statement

Edoardo Fazzari: Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft; **Donato Romano:** Supervision, Writing – review & editing; **Fabrizio Falchi:** Supervision, Writing – review & editing; **Cesare Stefanini:** Supervision, Writing – review & editing.

Data availability

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

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.

Acknowledgment

This work was partially carried out in the framework of the H2020 FETOPEN Project “Robocoenosis-ROBOts in cooperation with a bioCOENOSIS”899520]. The founder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Akçay, H. G., Kabasakal, B., Aksu, D., Demir, N., Öz, M., & Erdoğan, A. (2020). Automated bird counting with deep learning for regional bird distribution mapping. *Animals*, *10*, 1207.

Al-Thani, N., Albuainain, A., Alnaimi, F., & Zorba, N. (2020). Drones for sheep livestock monitoring. IEEE. In *2020 IEEE 20th mediterranean electrotechnical conference (MELECON)*, pp. 672–676

Alagele, M., & Yildirim, R. (2022). Animal gait identification using a deep learning method. IEEE. In *2022 International symposium on multidisciplinary studies and innovative technologies (ISMSIT)*, pp. 540–542.

Alameer, A., Buijs, S., O’Connell, N., Dalton, L., Larsen, M., Pedersen, L., & Kyriazakis, I. (2022). Automated detection and quantification of contact behaviour in pigs using deep learning. *Biosystems Engineering*, *224*, 118–130.

Alameer, A., Kyriazakis, I., Dalton, H. A., Miller, A. L., & Bacardit, J. (2020). Automatic recognition of feeding and foraging behaviour in pigs using deep learning. *Biosystems Engineering*, *197*, 91–104.

Andresen, N., Wöllhaf, M., Hohlbaum, K., Lewejohann, L., Hellwich, O., Thöne-Reineke, C., & Belik, V. (2020). Towards a fully automated surveillance of well-being status in laboratory mice using deep learning: Starting with facial expression analysis. *PloS One*, *15*, e0228059.

Arablouei, R., Wang, L., Currie, L., Yates, J., Alvarenga, F. A., & Bishop-Hurley, G. J. (2023a). Animal behavior classification via deep learning on embedded systems. *Computers and Electronics in Agriculture*, *207*, 107707.

Arablouei, R., Wang, Z., Bishop-Hurley, G. J., & Liu, J. (2023b). Multimodal sensor data fusion for in-situ classification of animal behavior using accelerometry and gnss data. *Smart Agricultural Technology*, *4*, 100163.

Ayadi, S., Ben Said, A., Jabbar, R., Aloulou, C., Chabbouh, A., & Achballah, A. B. (2020). Dairy cow rumination detection: A deep learning approach. Springer. In *Distributed computing for emerging smart networks: Second international workshop, DiCES-N 2020, Bizerte, Tunisia, December 18, 2020, Proceedings 2*, pp. 123–139

Banerjee, S. C., Khan, K. A., & Sharma, R. (2023). Deep-worm-tracker: Deep learning methods for accurate detection and tracking for behavioral studies in c. elegans. *Applied Animal Behaviour Science*, *266*, 106024.

Bao, J., & Xie, Q. (2022). Artificial intelligence in animal farming: A systematic literature review. *Journal of Cleaner Production*, *331*, 129956.

Barnard, S., Calderara, S., Pistocchi, S., Cucchiara, R., Podaliri-Vulpiani, M., Messori, S., & Ferri, N. (2016). Quick, accurate, smart: 3d computer vision technology helps assessing confined animals’ behaviour. *PloS One*, *11*, e0158748.

Benaissa, S., Tuyttens, F., Plets, D., Martens, L., Vandaele, L., Joseph, W., & Sonck, B. (2023). Improved cattle behaviour monitoring by combining ultra-wideband location and accelerometer data. *Animal : An International Journal of Animal Bioscience*, *17*, 100730.

Bergen, S., Huso, M. M., Duerr, A. E., Braham, M. A., Katzner, T. E., Schmuecker, S., & Miller, T. A. (2022). Classifying behavior from short-interval biologging data: An example with gps tracking of birds. *Ecology and Evolution*, *12*, e08395.

Berman, G. J., Choi, D. M., Bialek, W., & Shaevitz, J. W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, *11*, 20140672.

Bernardes, R. C., Lima, M. A. P., Guedes, R. N. C., da Silva, C. B., & Martins, G. F. (2021). Ethoflow: Computer vision and artificial intelligence-based software for automatic behavior analysis. *Sensors*, *21*, 3237.

Bhattacharya, S., & Shahnawaz, S. (2021). Pose recognition in the wild: Animal pose estimation using agglomerative clustering and contrastive learning. arXiv preprint arXiv:2111.08259.

Biderman, D., Whiteway, M. R., Hurwitz, C., Greenspan, N., Lee, R. S., Vishnubhotla, A., Warren, R., Pedraja, F., Noone, D., & Schartner, M., et al. (2023). Lightning pose: Improved animal pose estimation via semi-supervised learning, bayesian ensembling, and cloud-native open-source tools. *bioRxiv*.

Bjerge, K., Alison, J., Dyrmann, M., Frigaard, C. E., Mann, H. M., & Høye, T. T. (2023). Accurate detection and identification of insects from camera trap images with deep learning. *PLOS Sustainability and Transformation*, *2*, e0000051.

Blau, A., Gebhardt, C., Bendesky, A., Paninski, L., & Wu, A. (2022). Semimultipose: A semi-supervised multi-animal pose estimation framework. arXiv preprint arXiv:2204.07072.

Bocaj, E., Uzunidis, D., Kasnesis, P., & Patrikakis, C. Z. (2020). On the benefits of deep convolutional neural networks on animal activity recognition. IEEE. In *2020 International conference on smart systems and technologies (SST)*, pp. 83–88

Bohnslav, J. P., Wimalasena, N. K., Clausing, K. J., Dai, Y. Y., Yarmolinsky, D. A., Cruz, T., Kashlan, A. D., Chiappe, M. E., Orefice, L. L., & Woolf, C. J., et al. (2021). Deepethogram, a machine learning pipeline for supervised behavior classification from raw pixels. *ELife*, *10*, e63377.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on computational learning theory* (pp. 144–152).

Bravo Sanchez, F. J., Hossain, M. R., English, N. B., & Moore, S. T. (2021). Bioacoustic classification of avian calls from raw sound waveforms with an open-source deep learning architecture. *Scientific Reports*, *11*, 15733.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Brown, A. E., & de Bivort, B. (2017). The study of animal behaviour as a physical science. *bioRxiv*. https://doi.org/10.1101/220855,

Bryda, E. C. (2013). The mighty mouse: The impact of rodents on advances in biomedical research. *Missouri Medicine*, *110*, 207.

Burgos-Artizzu, X. P., Dollár, P., Lin, D., Anderson, D. J., & Perona, P. (2012). Social behavior recognition in continuous video. IEEE. In *2012 IEEE conference on computer vision and pattern recognition*, pp. 1322–1329

Cao, J., Tang, H., Fang, H. S., Shen, X., Lu, C., & Tai, Y. W. (2019). Cross-domain adaptation for animal pose estimation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9498–9507).

Cao, S., Zhao, D., Liu, X., & Sun, Y. (2020). Real-time robust detector for underwater live crabs based on deep learning. *Computers and Electronics in Agriculture*, *172*, 105339.

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7291–7299).

Chalmers, C., Fergus, P., Wich, S., & Longmore, S. (2021). Modelling animal biodiversity using acoustic monitoring and deep learning. IEEE. In *2021 International joint conference on neural networks (IJCNN)*, pp. 1–7

Chambers, R. D., Yoder, N. C., Carson, A. B., Junge, C., Allen, D. E., Prescott, L. M., Bradley, S., Wymore, G., Lloyd, K., & Lyle, S. (2021). Deep learning classification of canine behavior using a single collar-mounted accelerometer: Real-world validation. *Animals*, *11*, 1549.

Chen, C., Zhu, W., & Norton, T. (2021). Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning. *Computers and Electronics in Agriculture*, *187*, 106255.

Chen, C., Zhu, W., Steibel, J., Siegford, J., Han, J., & Norton, T. (2020). Recognition of feeding behaviour of pigs and determination of feeding time of each pig by a video-based deep learning method. *Computers and Electronics in Agriculture*, *176*, 105642.

Chen, J., Hu, M., Coker, D. J., Berumen, M. L., Costelloe, B., Beery, S., Rohrbach, A., & Elhoseiny, M. (2023). Mammalnet: A large-scale video benchmark for mammal recognition and behavior understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 13052–13061).

Choi, T., Pyenson, B., Liebig, J., & Pavlic, T. P. (2022). Beyond tracking: Using deep learning to discover novel interactions in biological swarms. *Artificial Life and Robotics*, *27*, 393–400.

Choi, Y., Chae, H., Lee, J., Park, D., & Chung, Y. (2021). Cat monitoring and disease diagnosis system based on deep learning. *Journal of Korea Multimedia Society*, *24*, 233–244.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251–1258).

Coria-Avila, G. A., Pfaus, J. G., Orihuela, A., Domínguez-Oliva, A., José-Pérez, N., Hernández, L. A., & Mota-Rojas, D. (2022). The neurobiology of behavior and its applicability for animal welfare: A review. *Animals*, *12*, 928.

Coulibaly, S., Kamsu-Foguem, B., Kamissoko, D., & Traore, D. (2022). Explainable deep convolutional neural networks for insect pest recognition. *Journal of Cleaner Production*, *371*, 133638.

Dang, T. H., Dang, N. H., Tran, V. T., & Chung, W. Y. (2022). A lorawan-based smart sensor tag for cow behavior monitoring. IEEE. In *2022 IEEE sensors*, pp. 1–4

Dell, A. I., Bender, J. A., Branson, K., Couzin, I. D., de Polavieja, G. G., Noldus, L. P., Pérez-Escudero, A., Perona, P., Straw, A. D., & Wikelski, M., et al. (2014). Automated image-based tracking and its application in ecology. *Trends in Ecology & Evolution*, *29*, 417–428.

Ditria, E. M., Lopez-Marcano, S., Sievers, M., Jinks, E. L., Brown, C. J., & Connolly, R. M. (2020). Automating the analysis of fish abundance using object detection: Optimizing animal ecology with deep learning. *Frontiers in Marine Science*, (p. 429).

Duporge, I., Kholiavchenko, M., Harel, R., Wolf, S., Rubenstein, D., Crofoot, M., Berger-Wolf, T., Lee, S., Barreau, J., & Kline, J. (2024). Baboonland dataset: Tracking primates in the wild and automating behaviour recognition from drone videos. arXiv preprint arXiv:2405.17698.

Eerdekens, A., Deruyck, M., Fontaine, J., Martens, L., De Poorter, E., Plets, D., & Joseph, W. (2020). Resampling and data augmentation for equines’ behaviour classification based on wearable sensor accelerometer data using a convolutional neural network. IEEE. In *2020 International conference on omni-layer intelligent systems (COINS)*, pp. 1–6

Elbarrany, A. M., Mohialdin, A., & Atia, A. (2023). The use of pose estimation for abnormal behavior analysis in poultry farms. IEEE. In *2023 5th novel intelligent and leading emerging sciences conference (NILES)*, pp. 33–36

Evangelista, I. R. S., Concepcion, R., Palconit, M. G. B., Bandala, A. A., & Dadios, E. P. (2022). Yolov7 and deepsort for intelligent quail behavioral activities monitoring. IEEE. In *2022 IEEE 14th international conference on humanoid, nanotechnology, information technology, communication and control, environment, and management (HNICEM)*, pp. 1–5

Fang, C., Zhang, T., Zheng, H., Huang, J., & Cuan, K. (2021). Pose estimation and behavior classification of broiler chickens based on deep neural networks. *Computers and Electronics in Agriculture*, *180*, 105863.

Fazzari, E., Carrara, F., Falchi, F., Stefanini, C., & Romano, D., et al. (2023). A workflow for developing biohybrid intelligent sensing systems. In *Proceedings of the italia intelligenza artificiale - thematic workshops co-located with the 3rd CINI national lab AIIS conference on artificial intelligence (ital IA 2023)* (p. 555–560). CEUR Workshop Proceedings.

Fazzari, E., Carrara, F., Falchi, F., Stefanini, C., & Romano, D. (2024). Using ai to decode the behavioral responses of an insect to chemical stimuli: Towards machine-animal computational technologies. *International Journal of Machine Learning and Cybernetics*, *15*, 1985–1994.

Fazzari, E., Romano, D., Falchi, F., & Stefanini, C. (2025). Selective state models are what you need for animal action recognition. *Ecological Informatics*, *85*, 102955.

Feichtenhofer, C., Fan, H., Malik, J., & He, K. (2019). Slowfast networks for video recognition. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6202–6211).

Feighelstein, M., Ehrlich, Y., Naftaly, L., Alpin, M., Nadir, S., Shimshoni, I., Pinho, R. H., Luna, S. P., & Zamansky, A. (2023). Deep learning for video-based automated pain recognition in rabbits. *Scientific Reports*, *13*, 14679.

Fuentes, A., Yoon, S., Park, J., & Park, D. S. (2020). Deep learning-based hierarchical cattle behavior recognition with spatio-temporal information. *Computers and Electronics in Agriculture*, *177*, 105627.

Fujimori, S., Ishikawa, T., & Watanabe, H. (2020). Animal behavior classification using deeplabcut. IEEE. In *2020 IEEE 9th global conference on consumer electronics (GCCE)* , pp. 254–257.

Gnanasekar, S. T., Yanushkevich, S., Van den Hoogen, N. J., & Trang, T. (2022). Rodent tracking and abnormal behavior classification in live video using deep neural networks. IEEE. In *2022 IEEE symposium series on computational intelligence (SSCI)*, pp. 830–837

Gore, S. V., Kakodkar, R., Del Rosario Hernández, T., Edmister, S. T., & Creton, R. (2023). Zebrafish larvae position tracker (z-lap tracker): A high-throughput deep-learning behavioral approach for the identification of calcineurin pathway-modulating drugs using zebrafish larvae. *Scientific Reports*, *13*, 3174.

Gotanda, K. M., Farine, D. R., Kratochwil, C. F., Laskowski, K. L., & Montiglio, P. O. (2019). Animal behavior facilitates eco-evolutionary dynamics. arXiv preprint arXiv:1912.09505.

Graving, J. M., Chae, D., Naik, H., Li, L., Koger, B., Costelloe, B. R., & Couzin, I. D. (2019). Deepposekit, a software toolkit for fast and robust animal pose estimation using deep learning. *ELife*, *8*, e47994.

Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752.

Haalck, L., Mangan, M., Webb, B., & Risse, B. (2020). Towards image-based animal tracking in natural environments using a freely moving camera. *Journal of Neuroscience Methods*, *330*, 108455.

Han, F., Zhu, J., Liu, B., Zhang, B., & Xie, F. (2020). Fish shoals behavior detection based on convolutional neural network and spatiotemporal information. *IEEE Access*, *8*, 126907–126926.

Han, Y., Chen, K., Wang, Y., Liu, W., Wang, Z., Wang, X., Han, C., Liao, J., Huang, K., & Cai, S. (2024). Multi-animal 3d social pose estimation, identification and behaviour embedding with a few-shot learning framework. *Nature Machine Intelligence*, *6*(1), 48–61. Nature Publishing Group UK London

Hara, K., Kataoka, H., & Satoh, Y. (2018). Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6546–6555).

Hart, B. L. (2011). Behavioural defences in animals against pathogens and parasites: Parallels with the pillars of medicine in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 3406–3417.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961–2969).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. Springer. In *Computer vision–ECCV 2016: 14th European conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pp. 630–645.

Higaki, S., Matsui, Y., Miwa, M., Yamamura, T., Hojo, T., Yoshioka, K., Vang, A., Negreiro, A., & Dórea, J. R. R. (2024). Leveraging computer vision-based pose estimation technique in dairy cows for objective mobility analysis and scoring system. *Computers and Electronics in Agriculture*, *217*, 108573. Elsevier

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*, 1735–1780.

Hou, J., He, Y., Yang, H., Connor, T., Gao, J., Wang, Y., Zeng, Y., Zhang, J., Huang, J., & Zheng, B., et al. (2020). Identification of animal individuals using deep learning: A case study of giant panda. *Biological Conservation*, *242*, 108414.

Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., Webb, G. I., Idoumghar, L., Muller, P. A., & Petitjean, F. (2020). Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery*, *34*, 1936–1962.

Jaafra, Y., Laurent, J. L., Deruyver, A., & Naceur, M. S. (2019). Reinforcement learning for neural architecture search: A review. *Image and Vision Computing*, *89*, 57–66. Elsevier

Jia, Y., Li, S., Guo, X., Lei, B., Hu, J., Xu, X. H., & Zhang, W. (2022). Selfee, self-supervised features extraction of animal behaviors. *ELife*, *11*, e76218.

Jiang, L., Lee, C., Teotia, D., & Ostadabbas, S. (2022a). Animal pose estimation: A closer look at the state-of-the-art, existing gaps and opportunities. *Computer Vision and Image Understanding*, (p. 103483).

Jiang, M., Rao, Y., Zhang, J., & Shen, Y. (2020). Automatic behavior recognition of group-housed goats using deep learning. *Computers and Electronics in Agriculture*, *177*, 105706.

Jiang, Z., Liu, Z., Chen, L., Tong, L., Zhang, X., Lan, X., Crookes, D., Yang, M. H., & Zhou, H. (2022b). Detecting and tracking of multiple mice using part proposal networks. *IEEE Transactions on Neural Networks and Learning Systems*.

Jiang, Z., Zhou, F., Zhao, A., Li, X., Li, L., Tao, D., Li, X., & Zhou, H. (2021). Multi-view mouse social behaviour recognition with deep graphic model. *IEEE Transactions on Image Processing*, *30*, 5490–5504.

Jin, Z., Shu, H., Hu, T., Jiang, C., Yan, R., Qi, J., Wang, W., & Guo, L. (2024). Behavior classification and spatiotemporal analysis of grazing sheep using deep learning. *Computers and Electronics in Agriculture*, *220*, 108894. Elsevier

Joo, K. H., Duan, S., Weimer, S. L., & Teli, M. N. (2022). Birds' eye view: Measuring behavior and posture of chickens as a metric for their well-being. arXiv preprint arXiv:2205.00069.

Júnior, T. D. C., & Rieder, R. (2020). Automatic identification of insects from digital images: A survey. *Computers and Electronics in Agriculture*, *178*, 105784.

Kamminga, J. W., Le, D. V., Meijers, J. P., Bisby, H., Meratnia, N., & Havinga, P. J. (2018). Robust sensor-orientation-independent feature selection for animal activity recognition on collar tags. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*, 1–27.

Kamminga, J. W., Meratnia, N., & Havinga, P. J. (2019). Dataset: Horse movement data and analysis of its potential for activity recognition. In *Proceedings of the 2nd workshop on data acquisition to analysis* (pp. 22–25).

Kasnesis, P., Doulgerakis, V., Uzunidis, D., Kogias, D. G., Funcia, S. I., González, M. B., Giannousis, C., & Patrikakis, C. Z. (2022). Deep learning empowered wearable-based behavior recognition for search and rescue dogs. *Sensors*, *22*, 993.

Kaufmann, J., & Schering, A. (2014). Analysis of variance ANOVA. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118445112.stat06938,

Kavlak, A., Pastell, M., & Uimari, P. (2023). Disease detection in pigs based on feeding behaviour traits using machine learning. *Biosystems Engineering*, *226*, 132–143.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

Kleanthous, N., Hussain, A., Khan, W., Sneddon, J., & Liatsis, P. (2022a). Deep transfer learning in sheep activity recognition using accelerometer data. *Expert Systems with Applications*, *207*, 117925.

Kleanthous, N., Hussain, A. J., Khan, W., Sneddon, J., Al-Shamma'a, A., & Liatsis, P. (2022b). A survey of machine learning approaches in animal behaviour. *Neurocomputing*, *491*, 442–463.

Knight, E. C., & Bayne, E. M. (2019). Classification threshold and training data affect the quality and utility of focal species data processed with automated audio-recognition software. *Bioacoustics*, *28*, 539–554.

Koch, G., Zemel, R., & Salakhutdinov, R., et al. (2015). Siamese neural networks for one-shot image recognition. Lille. In *ICML deep learning workshop*,

Koger, B., Deshpande, A., Kerby, J. T., Graving, J. M., Costelloe, B. R., & Couzin, I. D. (2023). Quantifying the movement, behaviour and environmental context of group-living animals using drones and computer vision. *Journal of Animal Ecology*.

Korelidou, V., Simitzis, P., Massouras, T., & Gelasakis, A. I. (2024). Infrared thermography as a diagnostic tool for the assessment of mastitis in dairy ruminants. *Animals*, *14*(18), 2691. MDPI

Labuguen, R., Matsumoto, J., Negrete, S. B., Nishimaru, H., Nishijo, H., Takada, M., Go, Y., Inoue, K. i., & Shibata, T. (2021). Macaquepose: A novel "in the wild" macaque monkey pose dataset for markerless motion capture. *Frontiers in Behavioral Neuroscience*, *14*, 581154.

Landgraf, T., Gebhardt, G. H., Bierbach, D., Romanczuk, P., Musiolek, L., Hafner, V. V., & Krause, J. (2021). Animal-in-the-loop: Using interactive robotic conspecifics to study social behavior in animal groups. *Annual Review of Control, Robotics, and Autonomous Systems*, *4*, 487–507.

Lauer, J., Zhou, M., Ye, S., Menegas, W., Schneider, S., Nath, T., Rahman, M. M., Di Santo, V., Soberanes, D., & Feng, G., et al. (2022). Multi-animal pose estimation, identification and tracking with deeplabcut. *Nature Methods*, *19*, 496–504.

Lea, C., Vidal, R., Reiter, A., & Hager, G. D. (2016). Temporal convolutional networks: A unified approach to action segmentation. Springer. In *Computer vision–ECCV 2016 workshops: Amsterdam, The Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part III 14*, pp. 47–54,

Lecomte, C. G., Audet, J., Harnie, J., & Frigon, A. (2021). A validation of supervised deep learning for gait analysis in the cat. *Frontiers in Neuroinformatics*, *15*, 712623.

Lee, S., Waugh, B., O'Dell, G., Zhao, X., Yoo, W. S., & Kim, D. H. (2021). Predicting fruit fly behaviour using tolc device and deeplabcut. IEEE. in *2021 IEEE 21st international conference on bioinformatics and bioengineering (BIBE)*, pp. 1–6

Lei, Y., Dong, P., Guan, Y., Xiang, Y., Xie, M., Mu, J., Wang, Y., & Ni, Q. (2022). Postural behavior recognition of captive nocturnal animals based on deep learning: A case study of bengal slow loris. *Scientific Reports*, *12*, 7738.

Li, C., & Lee, G. H. (2023). Scarcenet: Animal pose estimation with scarce annotations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 17174–17183).

Li, D., Zhang, K., Li, Z., & Chen, Y. (2020). A spatiotemporal convolutional network for multi-behavior recognition of pigs. *Sensors*, *20*, 2381.

Li, J., Xu, C., Jiang, L., Xiao, Y., Deng, L., & Han, Z. (2019a). Detection and analysis of behavior trajectory for sea cucumbers based on deep learning. *IEEE Access*, *8*, 18832–18840.

Li, S., Li, J., Tang, H., Qian, R., & Lin, W. (2019b). Atrw: a benchmark for amur tiger re-identification in the wild. arXiv preprint arXiv:1906.05586.

Li, Y., Yang, G., Su, Z., Li, S., & Wang, Y. (2023). Human activity recognition based on multienvironment sensor data. *Information Fusion*, *91*, 47–63.

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117–2125).

Liu, D., Hou, J., Huang, S., Liu, J., He, Y., Zheng, B., Ning, J., & Zhang, J. (2023). LoTE-animal: A long time-span dataset for endangered animal behavior understanding. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 20064–20075).

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. IEEE. In *2008 eighth IEEE international conference on data mining*, pp. 413–422,

Liu, R., Xu, G., Jia, C., Ma, W., Wang, L., & Vosoughi, S. (2020a). Data boost: Text data augmentation through reinforcement learning guided conditional generation. arXiv preprint arXiv:2012.02952.

Liu, X., Yu, S. y., Flierman, N., Loyola, S., Kamermans, M., Hoogland, T. M., & De Zeeuw, C. I. (2020b). Optiflex: Video-based animal pose estimation using deep learning enhanced by optical flow. *BioRxiv, 2020–04*.

Liu, Y., Xu, H., Liu, D., & Wang, L. (2022). A digital twin-based sim-to-real transfer for deep reinforcement learning-enabled industrial robot grasping. *Robotics and Computer-Integrated Manufacturing*, *78*, 102365.

Lostanlen, V., Palmer, K., Knight, E., Clark, C., Klinck, H., Farnsworth, A., Wong, T., Cramer, J., & Bello, J. P. (2019). Long-distance detection of bioacoustic events with per-channel energy normalization. arXiv preprint arXiv:1911.00417.

Lu, W., Zhao, Y., Wang, J., Zheng, Z., Feng, L., & Tang, J. (2023). Mammalclub: An annotated wild mammal dataset for species recognition, individual identification, and behavior recognition. *Electronics*, *12*(21), 4506. MDPI

Luxem, K., Mocellin, P., Fuhrmann, F., Kürsch, J., Miller, S. R., Palop, J. J., Remy, S., & Bauer, P. (2022). Identifying behavioral structure from deep variational embeddings of animal motion. *Communications Biology*, *5*, 1267.

Mahmud, M. S., Zahid, A., Das, A. K., Muzammil, M., & Khan, M. U. (2021). A systematic literature review on deep learning applications for precision cattle farming. *Computers and Electronics in Agriculture*, *187*, 106313.

Manduca, G., Zeni, V., Moccia, S., Milano, B. A., Canale, A., Benelli, G., Stefanini, C., & Romano, D. (2023). Learning algorithms estimate pose and detect motor anomalies in flies exposed to minimal doses of a toxicant. *Iscience*, *26*.

Mankin, R., Hagstrum, D., Guo, M., Eliopoulos, P., & Njoroge, A. (2021). Automated applications of acoustics for stored product insect detection, monitoring, and management. *Insects*, *12*, 259.

Manoharan, D. S. (2020). Embedded imaging system based behavior analysis of dairy cow. *Journal of Electronics and Informatics*, *2*, 148–154.

Manriquez P, R., Kotz, S. A., Ravignani, A., & De Boer, B. (2024). Bioacoustic classification of a small dataset of mammalian vocalisations using deep learning. *Bioacoustics*, *33*(4), 354–371. Taylor & Francis.

Marks, M., Jin, Q., Sturman, O., von Ziegler, L., Kollmorgen, S., von der Behrens, W., Mante, V., Bohacek, J., & Yanik, M. F. (2022). Deep-learning-based identification, tracking, pose estimation and behaviour classification of interacting primates and mice in complex environments. *Nature Machine Intelligence*, *4*, 331–340.

Marshall, J. D., Li, T., Wu, J. H., & Dunn, T. W. (2022). Leaving flatland: Advances in 3d behavioral measurement. *Current Opinion in Neurobiology*, *73*, 102522.

Martin-Abadal, M., Ruiz-Frau, A., Hinz, H., & Gonzalez-Cid, Y. (2020). Jellytoring: Real-time jellyfish monitoring based on deep learning object detection. *Sensors*, *20*, 1708.

Mathis, A., Biasi, T., Schneider, S., Yuksekgonul, M., Rogers, B., Bethge, M., & Mathis, M. W. (2021). Pretraining boosts out-of-domain robustness for pose estimation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 1859–1868).

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). Deeplabcut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*, 1281–1289.

Mathis, M. W., & Mathis, A. (2020). Deep learning tools for the measurement of animal behavior in neuroscience. *Current Opinion in Neurobiology*, *60*, 1–11.

McIntosh, D., Marques, T. P., Albu, A. B., Rountree, R., & De Leo, F. (2020). Movement tracks for the automatic detection of fish behavior in videos. arXiv preprint arXiv:2011.14070.

McKenzie-Smith, G. C., Wolf, S. W., Ayroles, J. F., & Shaevitz, J. W. (2023). Capturing continuous, long timescale behavioral changes in drosophila melanogaster postural data. arXiv preprint arXiv:2309.04044.

Mekruksavanich, S., Jantawong, P., & Jitpattanakul, A. (2022). Resnet-based deep neural network using transfer learning for animal activity recognition. IEEE. In *2022 6th international conference on information technology (InCIT)*, pp. 445–449.

Mendoza, Q. A., Pordesimo, L., Neilsen, M., Armstrong, P., Campbell, J., & Mendoza, P. T. (2023). Application of machine learning for insect monitoring in grain facilities. *AI*, *4*, 348–360.

Mishra, S., & Sharma, S. K. (2023). Advanced contribution of iot in agricultural production for the development of smart livestock environments. *Internet of Things*, *22*, 100724.

Modlmeier, A. P., Colman, E., Hanks, E. M., Bringenberg, R., Bansal, S., & Hughes, D. P. (2019). Ant colonies maintain social homeostasis in the face of decreased density. *ELife*, *8*, e38473.

Mohialdin, A. M., Elbarrany, A. M., & Atia, A. (2023). Chicken behavior analysis for surveillance in poultry farms.

Mondal, A., Nag, S., Prada, J. M., Zhu, X., & Dutta, A. (2023). Actor-agnostic multi-label action recognition with multi-modal query. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 784–794).

Morfi, V., Bas, Y., Pamuła, H., Glotin, H., & Stowell, D. (2019). Nips4bplus: A richly annotated birdsong audio dataset. *PeerJ Computer Science*, *5*, e223.

Mori, K., Yamauchi, N., Wang, H., Sato, K., Toyoshima, Y., & Iino, Y. (2022). Probabilistic generative modeling and reinforcement learning extract the intrinsic features of animal behavior. *Neural Networks*, *145*, 107–120.

Moummad, I., Serizel, R., & Farrugia, N. (2023). Regularized contrastive pre-training for few-shot bioacoustic sound detection. arXiv preprint arXiv:2309.08971.

Nasiri, A., Amirivojdan, A., Zhao, Y., & Gan, H. (2023). Estimating the feeding time of individual broilers via convolutional neural network and image processing. *Animals*, *13*, 2428.

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using deeplabcut for 3d markerless pose estimation across species and behaviors. *Nature Protocols*, *14*, 2152–2176.

Neethirajan, S. (2020). The role of sensors, big data and machine learning in modern animal farming. *Sensing and Bio-Sensing Research*, *29*, 100367.

Ng, X. L., Ong, K. E., Zheng, Q., Ni, Y., Yeo, S. Y., & Liu, J. (2022). Animal kingdom: A large and diverse dataset for animal behavior understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 19023–19034).

Nguyen, T. D., & Kresovic, M. (2022). A survey of top-down approaches for human pose estimation. arXiv preprint arXiv:2202.02656.

Nilsson, S. R., Goodwin, N. L., Choong, J. J., Hwang, S., Wright, H. R., Norville, Z. C., Tong, X., Lin, D., Bentzley, B. S., & Eshel, N., et al. (2020). Simple behavioral analysis (simba)–an open source toolkit for computer classification of complex social behaviors in experimental animals. *BioRxiv, 2020–04*.

Nolasco, I., Singh, S., Morfi, V., Lostanlen, V., Strandburg-Peshkin, A., Vidaña-Vila, E., Gill, L., Pamuła, H., Whitehead, H., & Kiskin, I., et al. (2023). Learning to detect an animal sound from five examples. *Ecological Informatics*, *77*, 102258.

Odo, A., Muns, R., Boyle, L., & Kyriazakis, I. (2023). Video analysis using deep learning for automatic quantification of ear biting in pigs. *IEEE Access*.

Oestreich, W. K., Oliver, R. Y., Chapman, M. S., Go, M., & McKenna, M. F. (2024). Listening to animal behavior to understand changing ecosystems. *Trends in Ecology & Evolution*. Elsevier

Oliveira, D. A. B., Pereira, L. G. R., Bresolin, T., Ferreira, R. E. P., & Dorea, J. R. R. (2021). A review of deep learning algorithms for computer vision systems in livestock. *Livestock Science*, *253*, 104700.

Otsuka, R., Yoshimura, N., Tanigaki, K., Koyama, S., Mizutani, Y., Yoda, K., & Maekawa, T. (2024). Exploring deep learning techniques for wild animal behaviour classification using animal-borne accelerometers. *Methods in Ecology and Evolution*, *15*(4), 716–731. Wiley Online Library

Özdaş, M. B., Uysal, F., & Hardalaç, F. (2023). Classification of retinal diseases in optical coherence tomography images using artificial intelligence and firefly algorithm. *Diagnostics*, *13*, 433.

Pan, Z., Chen, H., Zhong, W., Wang, A., & Zheng, C. (2023). A cnn-based animal behavior recognition algorithm for wearable devices. *IEEE Sensors Journal*, *23*, 5156–5164.

Papandreou, G., Zhu, T., Chen, L. C., Gidaris, S., Tompson, J., & Murphy, K. (2018). Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model. In *Proceedings of the european conference on computer vision (ECCV)*.

Papaspyros, V., Escobedo, R., Alahi, A., Theraulaz, G., Sire, C., & Mondada, F. (2023). Predicting long-term collective animal behavior with deep learning. *bioRxiv, 2023–02*.

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023b). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual ACM symposium on user interface software and technology* (pp. 1–22).

Pedersen, W. A., McMillan, P. J., Kulstad, J. J., Leverenz, J. B., Craft, S., & Haynatzki, G. R. (2006). Rosiglitazone attenuates learning and memory deficits in tg2576 alzheimer mice. *Experimental Neurology*, *199*, 265–273.

Peng, X. B., Coumans, E., Zhang, T., Lee, T. W., Tan, J., & Levine, S. (2020). Learning agile robotic locomotion skills by imitating animals. arXiv preprint arXiv:2004.00784.

Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S. H., Murthy, M., & Shaevitz, J. W. (2019). Fast animal pose estimation using deep neural networks. *Nature Methods*, *16*, 117–125.

Pereira, T. D., Tabris, N., Matsliah, A., Turner, D. M., Li, J., Ravindranath, S., Papadoyannis, E. S., Normand, E., Deutsch, D. S., & Wang, Z. Y., et al. (2022). Sleap: A deep learning system for multi-animal pose tracking. *Nature Methods*, *19*, 486–495.

Perez, M., & Toler-Franklin, C. (2023). Cnn-based action recognition and pose estimation for classifying animal behavior from videos: A survey. arXiv preprint arXiv:2301.06187.

Pham, T. D. (2022). Classification of caenorhabditis elegans locomotion behaviors with eigenfeature-enhanced long short-term memory networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *20*, 206–216.

Piergiovanni, A., & Ryoo, M. (2019). Temporal gaussian mixture layer for videos. PMLR. In *International Conference on Machine learning*, pp. 5152–5161,

Pillai, R., Gupta, R., Sharma, N., & Bansal, R. K. (2023). A deep learning approach for detection and classification of ten species of monkeys. IEEE. In *2023 International conference on smart systems for applications in electrical sciences (ICSSES)*, pp. 1–6,

Rabiner, L., & Juang, B. (1986). An introduction to hidden markov models. *IEEE ASSP Magazine*, *3*, 4–16.

Rahman, S. A., Song, I., Leung, M., Lee, I., & Lee, K. (2014). Fast action recognition using negative space features. *Expert Systems with Applications*, *41*, 574–587. https://www.sciencedirect.com/science/article/pii/S0957417413005812, https://doi.org/10.1016/j.eswa.2013.07.082.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788).

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, *28*.

Riekert, M., Klein, A., Adrion, F., Hoffmann, C., & Gallmann, E. (2020). Automatically detecting pig position and posture by 2d camera imaging and deep learning. *Computers and Electronics in Agriculture*, *174*, 105391. https://doi.org/10.1016/j.compag.2020.105391.

Rijsdijk, J., Wu, L., Perin, G., & Picek, S. (2021). Reinforcement learning for hyperparameter tuning in deep learning-based side-channel analysis. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, (pp. 677–707).

Robinson, T. P., Wint, G. W., Conchedda, G., Van Boeckel, T. P., Ercoli, V., Palamara, E., Cinardi, G., D'Aietti, L., Hay, S. I., & Gilbert, M. (2014). Mapping the global distribution of livestock. *PloS ONE*, *9*, e96084.

Romano, D., & Stefanini, C. (2021). Unveiling social distancing mechanisms via a fish-robot hybrid interaction. *Biological Cybernetics*, *115*, 565–573.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Springer. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th International conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*, pp. 234–241,

Russello, H., van der Tol, R., & Kootstra, G. (2022). T-Leap: Occlusion-robust pose estimation of walking cows using temporal information. *Computers and Electronics in Agriculture*, *192*, 106559.

Saleh, A., Sheaves, M., Jerry, D., & Azghadi, M. R. (2022). Adaptive uncertainty distribution in deep learning for unsupervised underwater image enhancement. arXiv preprint arXiv:2212.08983.

Saleh, D., Ahmed, M., Zaafan, M., Farouk, Y., & Atia, A. (2023). A pharmacology toolkit for animal pose estimation, tracking and analysis. IEEE. In *2023 International mobile, intelligent, and ubiquitous computing conference (MIUCC)*, pp. 1–7.

Samsudin, W., Harizan, M., Ibrahim, M., Karim, R., & Ibrahim, W. (2022). Zebrafish larvae locomotor activity detection using convolutional neural network (cnn).

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510–4520).

Schall, E., Kaya, I. I., Debusschere, E., Devos, P., & Parcerisas, C. (2024). Deep learning in marine bioacoustics: A benchmark for baleen whale detection. *Remote Sensing in Ecology and Conservation*, *10*(5), 642–654. Wiley Online Library

Schneider, S., Lee, J. H., & Mathis, M. W. (2023). Learnable latent embeddings for joint behavioural and neural analysis. *Nature*, (pp. 1–9).

Segalin, C., Williams, J., Karigo, T., Hui, M., Zelikowsky, M., Sun, J. J., Perona, P., Anderson, D. J., & Kennedy, A. (2021). The mouse action recognition system (mars) software pipeline for automated analysis of social behaviors in mice. *ELife*, *10*, e63720.

Selvaraju, R. R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., & Batra, D. (2016). Grad-cam: Why did you say that? arXiv preprint arXiv:1611.07450.

Shamoun-Baranes, J., Burant, J. B., van Loon, E. E., Bouten, W., & Camphuysen, C. (2017). Short distance migrants travel as far as long distance migrants in lesser black-backed gulls larus fuscus. *Journal of Avian Biology*, *48*, 49–57.

Shaw, J. K., & Lahrman, S. (2023). The human–animal bond–a brief look at its richness and complexities. *Canine and Feline Behavior for Veterinary Technicians and Nurses*, (pp. 88–105).

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Stowell, D. (2022). Computational bioacoustics with deep learning: A review and roadmap. *PeerJ*, *10*, e13152.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818–2826).

Tassinari, P., Bovo, M., Benni, S., Franzoni, S., Poggi, M., Mammi, L. M. E., Mattoccia, S., Di Stefano, L., Bonora, F., & Barbaresi, A., et al. (2021). A computer vision approach based on deep learning for the detection of dairy cows in free stall barn. *Computers and Electronics in Agriculture*, *182*, 106030.

Teixeira, A. C., Ribeiro, J., Morais, R., Sousa, J. J., & Cunha, A. (2023). A systematic review on automatic insect detection using deep learning. *Agriculture*, *13*, 713.

Temenos, A., Voulodimos, A., Korelidou, V., Gelasakis, A., Kalogeras, D., Doulamis, A., & Doulamis, N. (2024). Goat-CNN: A lightweight convolutional neural network for pose-independent body condition score estimation in goats. *Journal of Agriculture and Food Research*, *16*, 101174. Elsevier

Thanh, V. Q., & Netramai, C. (2022). Deep learning-based monitoring system for distress on mice using behavior analysis. IEEE. In *2022 International electrical engineering congress (iEECON)*, pp. 1–4,

Tian, Z., Chen, H., & Shen, C. (2019). Directpose: Direct end-to-end multi-person pose estimation. arXiv preprint arXiv:1911.07451.

Tilak, O., Mukhopadhyay, S., Tuceryan, M., & Raje, R. (2010). A novel reinforcement learning framework for sensor subset selection. IEEE. In *2010 International conference on networking, sensing and control (ICNSC)*, pp. 95–100.

Tucker Edmister, S., Del Rosario Hernández, T., Ibrahim, R., Brown, C. A., Gore, S. V., Kakodkar, R., Kreiling, J. A., & Creton, R. (2022). Novel use of fda-approved drugs identified by cluster analysis of behavioral profiles. *Scientific Reports*, *12*, 6120.

Uchino, T., & Ohwada, H. (2021). Individual identification model and method for estimating social rank among herd of dairy cows using yolov5. IEEE. In *2021 IEEE 20th international conference on cognitive informatics & cognitive computing (ICCI* CC)*, pp. 235–241,

Ullah, N., Khan, J. A., Alharbi, L. A., Raza, A., Khan, W., & Ahmad, I. (2022). An efficient approach for crops pests recognition and classification based on novel deeppestnet deep learning model. *IEEE Access*, *10*, 73019–73032.

Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, *124*, 203–220. Elsevier

Varma, A. L. S., Bateshwar, V., Rathi, A., & Singh, A. (2021). Acoustic classification of insects using signal processing and deep learning approaches. IEEE. In *2021 8th international conference on signal processing and integrated networks (SPIN)*, pp. 1048–1052,

Wang, J., Wang, N., Li, L., & Ren, Z. (2020). Real-time behavior detection and judgment of egg breeders based on yolo v3. *Neural Computing and Applications*, *32*, 5471–5481.

Wang, K., Wu, P., Cui, H., Xuan, C., & Su, H. (2021a). Identification and classification for sheep foraging behavior based on acoustic signal and deep learning. *Computers and Electronics in Agriculture*, *187*, 106275.

Wang, X., Du, C., Wang, Y., Hu, S., & Zhao, Y. (2021b). Behavioral recognition of mice based on a deep network. IEEE. In *2021 IEEE 11th annual international conference on CYBER technology in automation, control, and intelligent systems (CYBER)*, pp. 840–844.

Weber, R. Z., Mulders, G., Kaiser, J., Tackenberg, C., & Rust, R. (2022). Deep learning-based behavioral profiling of rodent stroke recovery. *BMC Biology*, *20*, 1–19.

Wijeyakulasuriya, D. A., Eisenhauer, E. W., Shaby, B. A., & Hanks, E. M. (2020). Machine learning for modeling animal movement. *PloS ONE*, *15*, e0235750.

Wittek, N., Wittek, K., Keibel, C., & Güntürkün, O. (2023). Supervised machine learning aided behavior classification in pigeons. *Behavior Research Methods*, *55*, 1624–1640.

Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. IEEE. In *2017 IEEE international conference on image processing (ICIP)*, pp. 3645–3649.

Wu, A., Buchanan, E. K., Whiteway, M., Schartner, M., Meijer, G., Noel, J. P., Rodriguez, E., Everett, C., Norovich, A., & Schaffer, E., et al. (2020). Deep graph pose: A semi-supervised deep graphical model for improved animal pose tracking. *Advances in Neural Information Processing Systems*, *33*, 6040–6052.

Wu, J., Leng, C., Wang, Y., Hu, Q., & Cheng, J. (2016). Quantized convolutional neural networks for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4820–4828).

Xiao, S., Wang, Y., Perkes, A., Pfrommer, B., Schmidt, M., Daniilidis, K., & Badger, M. (2023). Multi-view tracking, re-id, and social network analysis of a flock of visually similar birds in an outdoor aviary. *International Journal of Computer Vision*, *131*, 1532–1549.

Xu, W., Zhang, X., Yao, L., Xue, W., & Wei, B. (2020). A multi-view cnn-based acoustic classification system for automatic animal species identification. *Ad Hoc Networks*, *102*, 102115.

Xudong, Z., Xi, K., Ningning, F., & Gang, L. (2020). Automatic recognition of dairy cow mastitis from thermal images by a deep learning detector. *Computers and Electronics in Agriculture*, *178*, 105754.

Yamada, J., Shawe-Taylor, J., & Fountas, Z. (2020). Evolution of a complex predator-prey ecosystem on large-scale multi-agent deep reinforcement learning. IEEE. In *2020 International joint conference on neural networks (IJCNN)*, pp. 1–8,

Yamaguchi, S., Naoki, H., Ikeda, M., Tsukada, Y., Nakano, S., Mori, I., & Ishii, S. (2018). Identification of animal behavioral strategies by inverse reinforcement learning. *PLoS Computational Biology*, *14*, e1006122.

Yang, Y., Yang, J., Xu, Y., Zhang, J., Lan, L., & Tao, D. (2022). Apt-36k: A large-scale benchmark for animal pose estimation and tracking. *Advances in Neural Information Processing Systems*, *35*, 17301–17313.

Ye, S., Filippova, A., Lauer, J., Vidal, M., Schneider, S., Qiu, T., Mathis, A., & Mathis, M. W. (2022). Superanimal models pretrained for plug-and-play analysis of animal behavior. arXiv preprint arXiv:2203.07436.

Yu, H., Xu, Y., Zhang, J., Zhao, W., Guan, Z., & Tao, D. (2021). Ap-10k: A benchmark for animal pose estimation in the wild. arXiv preprint arXiv:2108.12617.

Zhang, K., Li, D., Huang, J., & Chen, Y. (2020). Automated video behavior recognition of pigs using two-stream convolutional networks. *Sensors*, *20*, 1085.

Zhou, F., Yang, X., Chen, F., Chen, L., Jiang, Z., Zhu, H., Heckel, R., Wang, H., Fei, M., & Zhou, H. (2022). Cross-skeleton interaction graph aggregation network for representation learning of mouse social behaviour. arXiv preprint arXiv:2208.03819.

Zhu, Y., Lan, Z., Newsam, S., & Hauptmann, A. (2019). Hidden two-stream convolutional networks for action recognition. In *Computer vision–ACCV 2018: 14th Asian conference on computer vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14*, Springer. pp. 363–378,