Individual Behavior Recognition in Laboratory Rats: a Comparative Analysis of Computer Vision Methods

Dmitrii Krasnov[1*], Aleksey Shmonin[2,3], Maxim Volynsky[1], Nikita Margaryants[1,4], Alexandr Gusev[1], Maria Maltseva[3], and Elena Korotkova[3]
[1] Technical Vision Lab, ITMO University, 49 Kronverksky pr., Saint-Petersburg 197101, Russian Federation
[2] Department of Psychology, Pedagogy and Rehabilitation Technologies, Federal Scientific and Clinical Center for Resuscitation and Rehabilitation, 777, build. 1 Village Lytkino, Moscow region, Solnechnogorsk 141534, Russian Federation
[3] Department of Pathophysiology with a course in Clinical Pathophysiology, Pavlov First Saint Petersburg State Medical University, 6–8 L'va Tolstogo str., Saint-Petersburg 197022, Russian Federation
[4] Department of Physics, Math, and Informatics, Pavlov First Saint Petersburg State Medical University, 6–8 L'va Tolstogo str., Saint-Petersburg 197022, Russian Federation
*e-mail: dmitriy_krasnov@outlook.com

Abstract. Computer-vision methods have been applied to automated behavior recognition in laboratory rodents. Namely, we explored the possibility of classifying certain behavior classes from still images; compared keypoint-based methods with approaches based on visual embeddings; studied the feasibility of transferring models between rats and mice; and evaluated the relevance of the number and the accuracy of the detected keypoints. We collected a dataset of a freely moving Wistar rat to train six pose-based classifiers with Long Short-Term Memory (LSTM) using six sets of keypoints produced by two detectors and four Convolutional Neural Network (CNN) classifiers using images and optical flow frames. The results demonstrated the highest mean average precision (mAP) of 65.7% for the CNN-based methods and 34.3% for the LSTM classifiers, the feasibility of recognizing visually distinct classes (rearing and body grooming) from still images, and the applicability of a keypoint detector trained on mice. The results of this study can be applied to the design of a computer vision system for automating long-term monitoring of laboratory rodent behavior.

Keywords: computer vision; deep learning; machine learning; rat behavior; rodent behavior; action recognition; keypoints; convolutional neural network; Wistar rat.

Paper #9476 received 15 Jan 2026; accepted for publication 19 Feb 2026; published online 25 Mar 2026. doi: 10.18287/JBPE26.12.010306.

1 Introduction

Analysis of animal behavior through video monitoring has become an important new task in neuroscience [1], since pharmaceuticals, physical environment, living conditions, diseases, and injuries tend to affect animal behavior. Such influence is typically evaluated via behavioral tests, including tests for isolated assessment of a specific skill and long-term monitoring.

Tests for the isolated assessment of a skill or specific behavioral acts, such as stress, anxiety, depression, or learned helplessness, are widely used in animal behavioral research [2]. In these tests, animals are placed in specific conditions for a short period, usually a few minutes. Specific behavioral forms are recorded, based on which conclusions about the presence or absence of target symptoms and deviations are drawn. Such analysis is possible in a "manual mode", where the researcher manually counts the required behavioral acts [2].

Long-term monitoring of animals – for several hours – is necessary for assessing rare behavioral activities, observing general locomotor activity (walking and running distance, duration of sitting and sleeping), and monitoring interaction between males and females [3]. As long-term monitoring requires analyzing hours of video and often involves a behavior analysis expert, it necessitates automation via modern computer vision methods [3].

Long-term monitoring allows for the assessment of disability in experiments. Disability is a limitation of an individual's ability to participate in normal life and perform everyday tasks. An individual with a disability does not use certain behavioral patterns, for example sleeps a lot, moves less, does not communicate, drinks and eats less or more. Analyzing the quantity, quality, and sequence of actions, can solve the following tasks:

- Creating a disability profile in an experiment, where the researcher can analyze which behavioral forms are impaired in each disease and prescribe drugs, rehabilitation technologies, and environmental intervention options to restore the individual's previous activity level, in other words, to eliminate the disability;
- Monitoring an animal in a cage at home or in a zoo to detect signs of pathology based on behavior and provide necessary recommendations [4];
- Investigating psychotropic drugs that can alter animal behavior to evaluate long-term effects and changes in individual and social behavior.

Rodents, such as rats and mice, are appropriate candidates for behavioral experiments, since they are social animals living in families, on which various diseases can be modeled using pharmacological, genetic, and surgical experimental models.

Currently, there are many approaches for recognizing the behavior patterns of laboratory rodents, based on spatiotemporal information obtained from video sequences [5, 6]. Most of them are based on analyzing features extracted from key points, reflecting their spatial configuration and change over time. Predictions can be made using either heuristic methods [7] or a trained classifier [8]. However, this approach requires a significant amount of data not only for training the classifier but also for training the keypoint detector, as the robustness and generalization capability of the entire pipeline depend on it [9].

Another drawback of this approach is the narrow specialization of existing open-source and commercial solutions. This can manifest in a specific type of animal (mouse or rat), color (white, black, gray, brown), or number of species. For example, Segalin et al. [10] presents a framework for recognizing social interactions between a white and a black mouse. Hsu and Yttri [11] show a pipeline for monitoring the behavior of a spotted dark rat. Differences can also be in camera angles (top-down or bottom-up) or in the chosen keypoints for describing the rodent skeleton. Such variability can lead to difficulties in scaling the particular approach to other animals, for example Wistar rats [12].

A further review of the literature revealed that existing studies have not examined how the complexity of the skeletal model affects classification accuracy when evaluated on the same data. This is an important aspect of pose-based methods, as certain skeletal keypoints (hind limbs and forelimbs) are typically detected less accurately than others (nose, tail base, spine) due to more active motion and frequent occlusions. However, these keypoints may be critical for recognizing complex behaviors such as scratching, face grooming, and body grooming.

Another unexplored question is the potential for directly using open-source mouse datasets to address rat behavior recognition tasks. Despite the visible similarity between rats and mice, they differ in body size and shape, as well as in individual and social behaviors. Moreover, the color of animals in open-source datasets may differ substantially from that of individuals in the target task, which can also affect behavior recognition accuracy. Thus, open-source datasets may be effectively used to train robust species and keypoint detectors with strong generalization capability due to their large size and high variability.

At the same time, methods for human action recognition commonly include convolutional neural networks (CNN) with Conv3D layers [13], hybrid architectures for processing frames and optical flow [14], or visual-language models [15]. It can be assumed that methods that successfully classify human activities are capable of recognizing the behavior of laboratory rodents. However, relatively few existing studies focused on adapting human action-recognition methods to this task. Notable examples include studies [16, 17], in which the authors used YOLO-family convolutional networks for localizing and classifying individual behaviors, as well as work [18], where visual embeddings extracted with I3D [19] were used for classification. However, none of these studies performed a direct comparison between pose-based and CNN-based methods.

Thus, the aim of this work is to investigate several important aspects of applying computer-vision methods to automate behavioral analysis in Wistar rats. Namely, the possibility of classifying certain behavior classes from still images; the comparison keypoint-based methods with approaches based on visual embeddings; the feasibility of transferring models between rats and mice; and the relevance of the number and the accuracy of the detected keypoints.

2 Methods and Materials

2.1 Animal Housing

The experiments used healthy Wistar rats weighing 250–350 g from the "Rappolovo" nursery breeding (Leningrad region, Russian Federation). The animals were housed in a 12-h light/12-h dark cycle with 5 rats per cage and free access to food and water. Behavioral analysis was conducted in the daytime under ambient lighting. During video recording the rats were housed individually with free access to food and water, in the behavioral testing boxes with wood shavings and no researchers in the room. The procedures were performed in accordance with the Guide for the Care and Use of Laboratory Animals and approved by the local Ethics Committee.

2.2 Data Collection and Processing

The video was recorded using an IDIS DC-Z1163 camera (IDIS, South Korea) with 1920×1080 resolution and 30 frames per second frame rate. The camera was positioned above two transparent plastic boxes measuring 120×60×40 cm³ at a distance of 3 m from the box lid. This setup resulted in one box occupying an area of approximately 1060×540 pixels in the frame, with a rat occupying about 5% of the box's area.

Two datasets were compiled and annotated based on video recordings of rat behavior, and the third one was collected from open sources. The first dataset consisted of 3400 images labelled with keypoints that were approved by experts in rat behavior from the Department of Pathophysiology, First Pavlov State Medical University of St. Petersburg (Russia). Fig. 1 shows an example of the annotated image: (1) nose; (2 and 3) right and left eye; (5 and 6) right and left ear; (4) head center; (7) forelimb girdle center; (8) spine center; (9 and 10) right and left forelimb; (11) tail base; (12 and 13) right and left hind limb [20, 21].

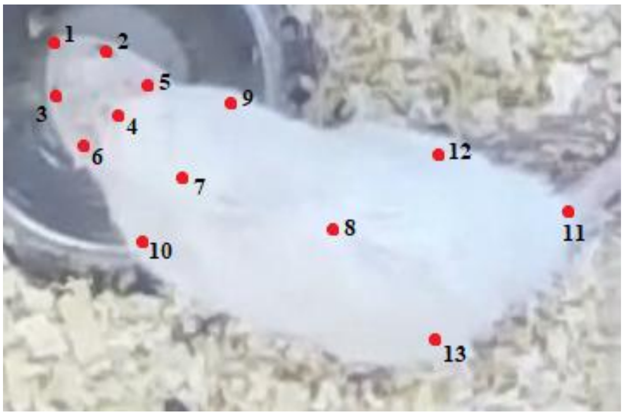

Fig. 1 Annotated image.

The second dataset consisted of video recordings of rat behavior patterns, where each recording was assigned to one of six classes: sniffing, rearing, eating, body grooming, face grooming, and scratching. The dataset contained 158 video recordings with a total duration of 43 min. A qualified behavior expert selected the specified behavioral forms and trimmed the videos from the start to the completion of the behavioral act.

To analyze the scalability of the approaches across the species of laboratory rodents, a dataset from open sources [10, 22] was also compiled. This dataset contained 14000 images of laboratory mice with annotated skeletal keypoints: nose, left and right ear, neck, spine center, left and right hip, tail base.

2.3 Deep Learning Approaches

For comparison, behavior patterns were recognized with two main approaches: a neural network classifier based on keypoint features and CNNs for processing raw frames and optical flow frames.

A long short-term memory (LSTM) block (two layers, sequence length 32, 256 hidden neurons) was used as the neural network classifier. A sequence of keypoint feature vectors containing the coordinates of selected points, mutual Euclidean distances and angles between them, and their displacements between adjacent frames was fed as the input for the classifier.

A single RGB frame and a sequence of optical flow frames were processed with the ResNet50 convolutional network [23]; for the RGB frames sequence, the version of ResNet18 with Conv3D layers [24] was employed. The optical flow frames were calculated using the dense Farneback algorithm [25] implemented in the OpenCV library.

To sum up, three CNNs were trained: for the RGB frame, for the sequence of 10 optical flow frames, and for the sequence of 32 RGB frames (Conv3D). Additionally, to analyze the possibility of fusing features from RGB frames and optical flow, an approach similar to the one in Ref. [14] was implemented, where a trainable 1×1 convolutional layer was used for the weighted sum of feature maps obtained from the frozen CNNs previously trained on RGB frames and optical flow frames.

2.4 Design of Experiments

The first approach was evaluated using 6 sets of keypoints in the experiments with 2 the above datasets (Fig. 2):

1) Open-source dataset: nose, neck, spine center, tail base;

2) Open-source dataset: nose, right and left ear, neck, spine center, tail base;

3) Open-source dataset: nose, right and left ear, neck, spine center, right and left hip, tail base;

4) Our dataset: nose, head center, forelimb girdle center, spine center, tail base;

5) Our dataset: nose, right and left eye, head center, right and left ear, forelimb girdle center, spine center, tail base;

6) Our dataset: nose, right and left eye, head center, right and left ear, neck, spine center, right and left forelimb, tail base, right and left hind limb.

The main idea behind this design was to assess the impact of the number and detection accuracy of keypoints on the accuracy of behavior class determination, since points along the spine and on the head are detected with higher accuracy [20].

For this purpose, an RTMDet-s [26] rat detector was trained for the spatial localization of the rat, and two HRNet [27] skeletal keypoint detectors were trained for the open-source dataset and for our dataset, respectively. To train the LSTM classifiers for each of the 6 keypoint sets, the predictions from the rat detector and the keypoint detector for each video frame were collected and stored locally. This step accelerated the training process since keypoint coordinates for each video only needed to be computed once.

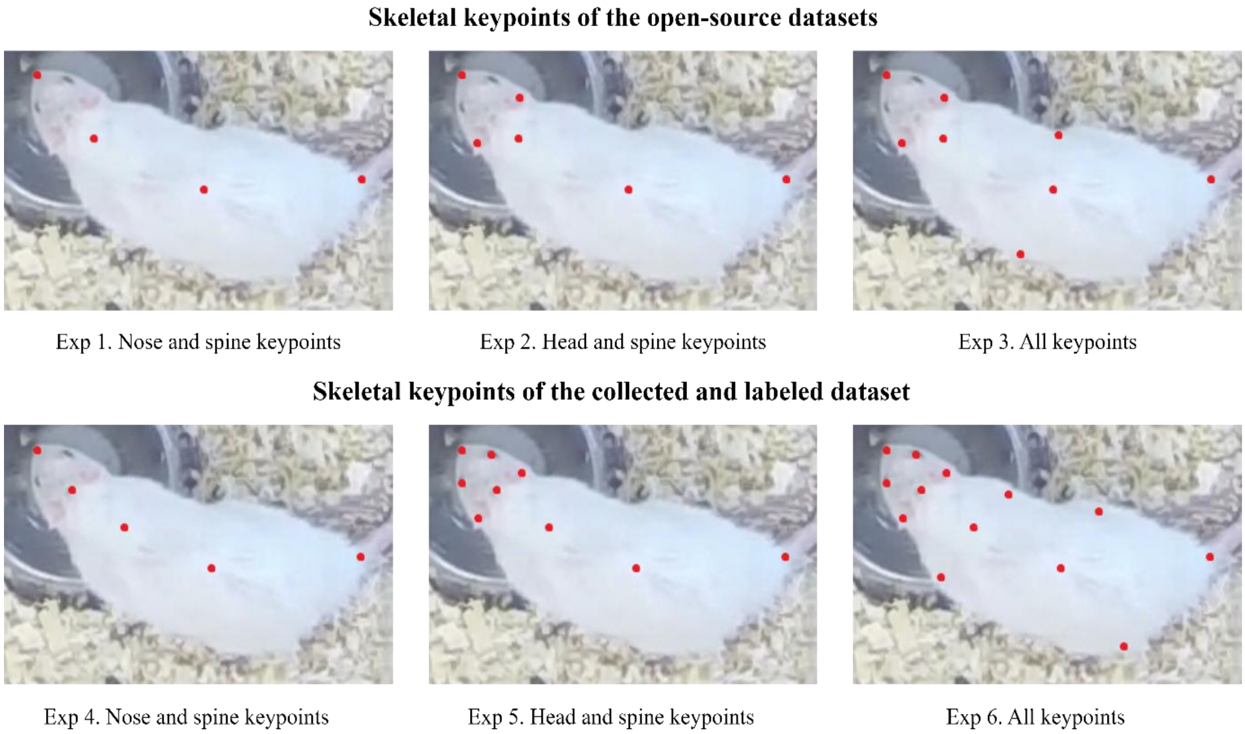

Fig. 2 Six sets of skeletal keypoints used in this work.

A similar step was performed during training the CNNs that require the use of optical flow. For each video, the optical flow was calculated using OpenCV, and its frames were saved for use in training. To evaluate the second approach, 4 CNNs were trained: for the RGB frame, for the sequence of 10 optical flow frames, for the sequence of 32 RGB frames (Conv3D), and for the fused RGB and optical flow features. All input frames were cropped according to the predicted rat bounding rectangle.

All models were trained on an NVIDIA GeForce RTX 3090 24GB for 100 epochs with a batch size of 32, Adam optimizer with learning rate of 0.00001. The models were evaluated using the Average Precision (AP) metric for each class and the mean value across all classes (mAP). The AP metric is calculated as the area under the Precision-Recall curve and indicates the overall quality of the classifier across all possible thresholds at which a class is considered correctly predicted.

3 Results and Discussion

This study examined some important aspects of recognizing individual behavior in laboratory rats: the applicability of human-oriented CNN-based action-recognition methods; the impact of both the number and the accuracy of the detected keypoints; and the transferability of detectors trained on mice. We trained six LSTM classifiers using different sets of keypoints produced by different detectors and four CNN classifiers. All models were trained on a dataset of video recordings of individual behavior in Wistar rats, that was collected in this work. The evaluation results are summarized in Table 1 as the AP for each class and the overall mAP.

Table 1 shows that the methods based on the skeletal keypoints (LSTM 1-6) scored as much as 47.8% lower in mAP than those based on the frame or a sequence of frames embeddings. This can be explained by the fact that the trained keypoint detector was not sufficiently robust for all observed rat poses. In contrast, CNN models do not require any complex or precise keypoint annotation, relying instead on the frame's visual features.

Among the CNN models, the fusion of RGB frame features and optical flow achieved the highest result (65.7% mAP), which agrees with the findings in Ref. [14]. We also observed that some behavior patterns, such as rearing or body grooming, were easier to be classified based on the single frame due to the distinct differences in the rat poses, which could be identified visually even in the still image. Moreover, the results demonstrated that the Conv2D-based ResNet achieved higher accuracy (62.8% mAP) than its Conv3D-based counterpart (61.1% mAP), which may indicate both that still-image embeddings were sufficient for successful action recognition and that temporal features were not used effectively in Conv3D layers. An absence of the distinct differences in the rat poses may also explain the lower metrics for the scratching, face grooming, sniffing, and eating classes, since the rat postures during these behaviors are very similar. Furthermore, the above classes accounted for a smaller total video duration. Although the CNNs trained in this work were previously reported as methods for recognizing human behavioral activities, our findings showed that they can also be applied to laboratory rodents with the more commonly used keypoint-based approaches.

Table 1 Classifiers evaluation results.

Experiment	AP for each class, %						mAP, %
	Rearing	Body grooming	Scratching	Face grooming	Sniffing	Eating	
LSTM Experiment 1	14.5	27.3	29.0	7.1	7.9	15.8	16.9
LSTM Experiment 2	34.9	29.4	17.4	21.1	13.1	40.4	26.0
LSTM Experiment 3	39.1	26.8	23.2	6.7	29.5	46.9	28.7
LSTM Experiment 4	12.6	51.3	40.9	7.5	17.4	13.3	23.8
LSTM Experiment 5	13.7	43.9	27.0	10.4	32.0	27.7	25.8
LSTM Experiment 6	30.1	58.0	26.8	29.8	33.1	28.0	34.3
ResNet50 RGB	82.2	76.1	59.6	58.8	46.6	53.6	62.8
ResNet18 Conv3D RGB	79.4	67.1	37.5	81.5	55.6	45.8	61.1
ResNet50 Optical Flow	67.9	71.5	68.1	43.6	40.3	62.6	59.0
ResNet50 RGB + Optical Flow	73.7	80.7	69.8	58.5	43.1	68.6	65.7

According to Table 1, the first group of LSTM experiments (1–3) tends to show similar results as the second group (4–6) – 28.7% and 34.3% mAP respectively. As mentioned above, 1–3 LSTM experiments used a keypoint detector trained on open-source datasets containing images mostly of mice, whereas 4–6 used the detector trained on the rat images collected in this work. These results suggest that trained detectors may be transferable across different rodent species (mice or rats) without any substantial loss in accuracy. Within each group, mAP increased as new keypoints were added (as the skeletal model became more detailed), even though the detector recognized the new points (right and left hip, right and left forelimb) with lower accuracy than the basic ones (nose, neck, spine center, tail base). This may indicate that a well-designed skeletal model played an important role in behavior-prediction accuracy, which slightly differs from the findings reported in Ref. [9].

Despite relying on a relatively old CNN backbone (ResNet, 2015) and a modest amount of data (a total video duration of 43 min), this study achieved metrics that are high enough for practical use (65.7% mAP). In addition, it may provide a valuable comparison of action recognition approaches in the perspective of classifying individual behaviors in Wistar rats, the results of which may be applied to long-term behavioral monitoring.

4 Conclusion

In this work, CNN-based and keypoint-based action-recognition approaches for identifying individual behaviors of Wistar rats were compared on a single dataset. The effects of the number of keypoints and the use of a detector trained on mice on classification accuracy were evaluated, and the feasibility of classifying behavior from a static crop using a CNN was assessed. The experimental results showed that CNNs were able to classify behavior from image embeddings with higher accuracy (65.7% mAP), and that classes such as rearing, body grooming, and scratching could be recognized with 61.1% mAP using a still image. For keypoint-based classifiers, the results indicated that a keypoint detector trained on mice images can be used without any substantial loss of accuracy for practical applications in rats (28.7% mAP and 34.3% mAP), and that a well-designed keypoint model is essential for robust behavior monitoring. This has clear implications for long-term behavior monitoring in laboratory rats, and suggests that publicly available keypoints datasets (mostly with mice images) can improve recognition accuracy in the conditions of small amounts of data, and that image embeddings are also a valuable source of information for this task.

Some questions remain unaddressed; for example, more recent feature extraction backbones – rather than only ResNet50 – could be explored, as well as some alternative methods for processing sequences of feature vectors, such as transformers or Conv1D CNNs. Since this study focused on a comparative analysis of approaches rather than on optimizing a specific pipeline, these aspects remain open for future investigation. Despite this, the study achieved the accuracy levels suitable for practical use (65.7% mAP). However, they could be improved by incorporating some novel methods in further work. Further research may also include expanding the dataset, adding new classes of individual behaviors such as nesting and drinking, and introducing social interaction to enable comprehensive behavioral analysis.

The results of this study can be applied to the design of a computer vision system for automating long-term monitoring of laboratory animal behavior. The findings demonstrate the importance of using image embeddings for behavior recognition, even when a powerful tool such as skeletal keypoints is available. The work may also help address the problem of the small amount of training data for keypoint detectors training and the challenges associated with annotating such datasets by suggesting the use of publicly available data in spite of the differences between rats and mice in behavioral and apparent aspects.

Disclosures

All authors declare that there is no conflict of interest in this paper.

Data and Code Availability

Data and code used in this work are available upon reasonable request from the corresponding author (Dmitrii Krasnov, dmitriy_krasnov@outlook.com).

References

1. F. Rotaru, S. I. Bejinariu, H. N. Costin, R. Luca, and C. D. Nita, "Captive animal behavior study by video analysis," Sensors 23(18), 7928 (2023).
2. Himanshu, Dharmila, D. Sarkar, and Nutan, "A review of behavioral tests to evaluate different types of anxiety and anti-anxiety effects," Clinical Psychopharmacology and Neuroscience: The Official Scientific Journal of the Korean College of Neuropsychopharmacology 18(3), 341–351 (2020).
3. F. Grieco, B. J. Bernstein, B. Biemans, et al., "Measuring behavior in the home cage: Study design, applications, challenges, and perspectives," Frontiers in Behavioral Neuroscience 15, 735387 (2021).
4. K. Luxem, J. J. Sun, S. P. Bradley, K. Krishnan, E. Yttri, J. Zimmermann, T. D. Pereira, and M. Laubach, "Open-source tools for behavioral video analysis: Setup, methods, and best practices," eLife 12, e79305 (2023).
5. M. Perez and C. Toler-Franklin, "CNN-based action recognition and pose estimation for classifying animal behavior from videos: A survey," arXiv:2301.06187 (2023).
6. S. İşık, G. Ünal, "Open-source software for automated rodent behavioral analysis," Frontiers in Neuroscience 17, 1149027 (2023).
7. C. J. Gabriel, Z. Zeidler, B. Jin, C. Guo, C. M. Goodpaster, A. Q. Kashay, A. Wu, M. Delaney, J. Cheung, L. E. DiFazio, M. J. Sharpe, D. Aharoni, S. A. Wilke, and L. A. DeNardo, "BehaviorDEPOT is a simple, flexible tool for automated behavioral detection based on markerless pose tracking," eLife 11, e74314 (2022).
8. M. Kabra, A. A. Robie, M. Rivera-Alba, S. Branson, and K. Branson, "JAABA: interactive machine learning for automatic annotation of animal behavior," Nature Methods 10(1), 64–67 (2013).
9. M. Lorbach, R. Poppe, E. A. Van Dam, L. P. J. J. Noldus, and R. C. Veltkamp, "Automated Recognition of Social Behavior in Rats: The Role of Feature Quality," in Image Analysis and Processing – ICIAP 2015, V. Murino and E. Puppo (Eds.), Springer International Publishing, 9280, 565–574 (2015).
10. C. Segalin, J. Williams, T. Karigo, M. Hui, M. Zelikowsky, J. J. Sun, P. Perona, D. J. Anderson, and A. Kennedy, "The Mouse Action Recognition System (MARS) software pipeline for automated analysis of social behaviors in mice," eLife 10, e63720 (2021).
11. A. I. Hsu, E. A. Yttri, "B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors," Nature Communications 12, 5188 (2021).
12. E. A. Van Dam, L. P. J. J. Noldus, and M. A. J. Van Gerven, "Deep learning improves automated rodent behavior recognition within a specific experimental setup," Journal of Neuroscience Methods 332, 108536 (2020).
13. J. Nie, L. Yan, X. Wang, and J. Chen, "A Novel 3D Convolutional Neural Network for Action Recognition in Infrared Videos," in 2021 4th International Conference on Information Communication and Signal Processing (ICICSP), IEEE, 420–424 (2021).
14. C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1933–1941 (2016).
15. Q. Wang, J. Du, K. Yan, and S. Ding, "Seeing in flowing: Adapting CLIP for action recognition with motion prompts learning," Proceedings of the 31st ACM International Conference on Multimedia (ACM), 5339–5347 (2023).
16. Z. Bai, Y. Li, Z. Ru, S. Li, W. Li, and F. Duan, "Rats Behavior Recognition Based on Improved YOLOv8," in 2024 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE, 1082–1086 (2024).
17. K. M. Hatton-Jones, C. Christie, T. A. Griffith, A. G. Smith, S. Naghipour, K. Robertson, J. S. Russell, J. N. Peart, J. P. Headrick, A. J. Cox, and E. F. Du Toit, "A YOLO based software for automated detection and analysis of rodent behaviour in the open field arena," Computers in Biology and Medicine 134, 104474 (2021).

18. N. G. Nguyen, D. Phan, F. R. Lumbanraja, M. R. Faisal, B. Abapihi, B. Purnama, M. K. Delimayanti, K. R. Mahmudah, M. Kubo, and K. Satou, "Applying Deep Learning Models to Mouse Behavior Recognition," Journal of Biomedical Science and Engineering 12(02), 183–196 (2019).
19. J. Carreira, A. Zisserman, "Quo Vadis, action recognition? A new model and the Kinetics dataset," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4724–4733 (2017).
20. D. I. Krasnov, M. A. Volynsky, and A. A. Gusev, "Preprocessing of skeletal keypoints trajectories in the task of laboratory animal behavior recording automation," Scientific and Technical Journal of Information Technologies, Mechanics and Optics 25(2), 295–302 (2025). [in Russian]
21. D. I. Krasnov, M. A. Volynsky, and A. A. Gusev, "State estimation of an agent with stochastic dynamics using recurrent filters," Informatics and Automation 24(5), 1355–1378 (2025). [in Russian]
22. N. L. Goodwin, J. J. Choong, S. Hwang, K. Pitts, L. Bloom, A. Islam, Y. Y. Zhang, E. R. Szelenyi, X. Tong, E. L. Newman, K. Miczek, H. R. Wright, R. J. McLaughlin, Z. C. Norville, N. Eshel, M. Heshmati, S. R. O. Nilsson, and S. A. Golden, "Simple Behavioral Analysis (SimBA) as a platform for explainable machine learning in behavioral neuroscience," Nature Neuroscience 27(7), 1411–1424 (2024).
23. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778 (2016).
24. K. Hara, H. Kataoka, and Y. Satoh, "Learning spatio-temporal features with 3D residual networks for action recognition," Proceedings of the IEEE International Conference on Computer Vision Workshop (ICCVW), 3154–3160 (2017).
25. G. Farnebäck, "Two-Frame Motion Estimation Based on Polynomial Expansion," in Image Analysis, J. Bigun and T. Gustavsson (Eds.), Springer Berlin Heidelberg, 2749, 363–370 (2003).
26. C. Lyu, W. Zhang, H. Huang, Y. Zhou, Y. Wang, Y. Liu, S. Zhang, and K. Chen, "RTMDet: An empirical study of designing real-time object detectors," arXiv:2212.07784 (2022).
27. J. Huang, Z. Zhu, F. Guo, and G. Huang, "The devil is in the details: Delving into unbiased data processing for human pose estimation," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5699–5708 (2020).