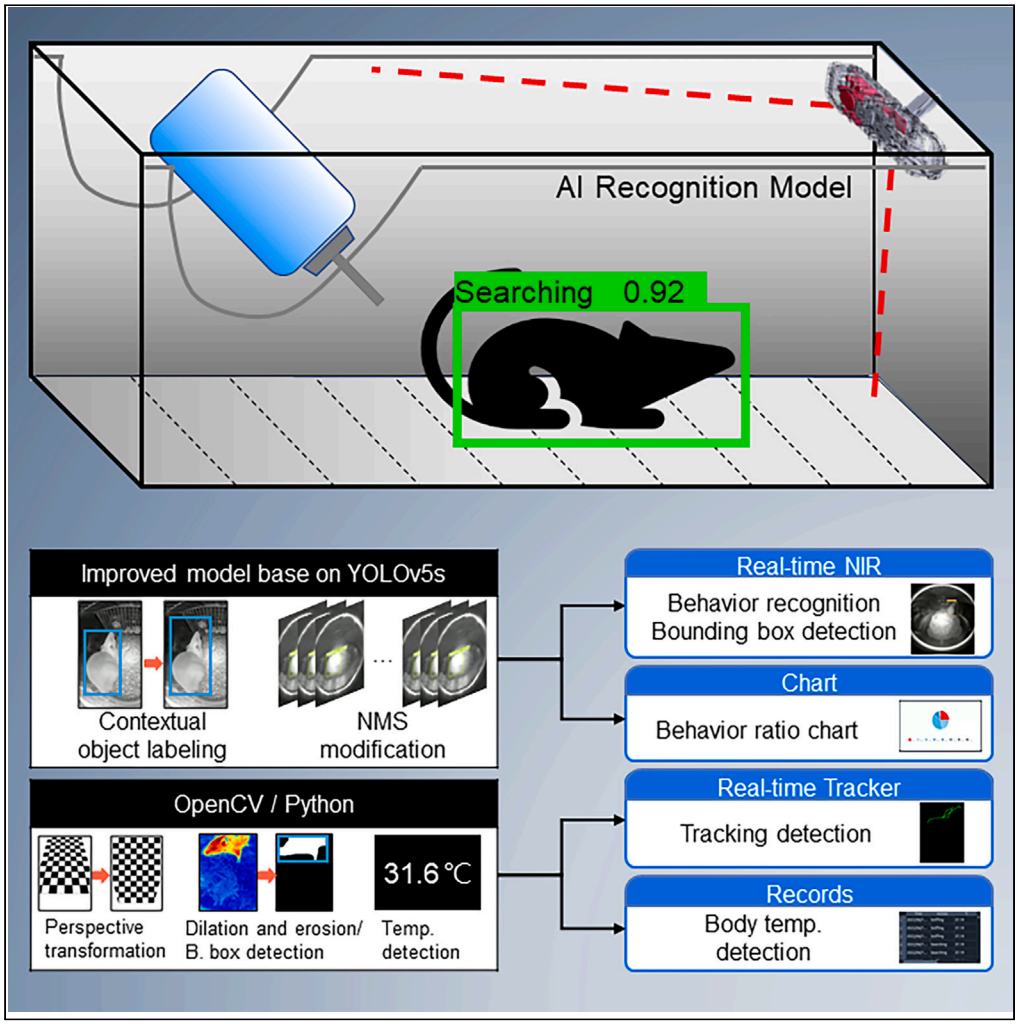
iScience

CellPress
OPEN ACCESS

Article

AI-powered home cage system for real-time tracking and analysis of rodent behavior



Chia-Ming Hsieh, Ching-Han Hsu, Jen-Kun Chen, Lun-De Liao

jkchen@nhri.edu.tw (J.-K.C.)
ldliao@nhri.edu.tw (L.-D.L.)

Highlights

A deep learning model is developed for recognizing rat behavior without sensors

Real-time behavior recognition, location tracking, and temperature detection

An F1 score exceeding 0.8 is achieved across five distinct behavioral categories

Designed for conventional cages, enabling home cage monitoring without relocation

Hsieh et al., iScience 27, 111223
November 15, 2024
https://doi.org/10.1016/j.isci.2024.111223

iScience

CellPress
OPEN ACCESS

Article

AI-powered home cage system for real-time tracking and analysis of rodent behavior

Chia-Ming Hsieh,[1,2] Ching-Han Hsu,[2] Jen-Kun Chen,[1,3,*] and Lun-De Liao[3,4,*]

SUMMARY

Researchers in animal behavior and neuroscience devote considerable time to observing rodents behavior and physiological responses, with AI monitoring systems reducing personnel workload. This study presents the RodentWatch (RW) system, which leverages deep learning to automatically identify experimental animal behaviors in home cage environments. A single multifunctional camera and edge device are installed inside the animal's home cage, allowing continuous real-time monitoring of the animal's behavior, position, and body temperature for extended periods. We investigated identifying the drinking and resting behaviors of rats, with recognition accuracy enhanced through contextual object labeling and modified non-maximum suppression (NMS) schemes. Two tests—a light cycle change test and a sucrose preference test—were conducted to evaluate the usability of this system in rat behavioral experiments. This system enables notable advancements in image-based behavior recognition for living rodents.

INTRODUCTION

The study of experimental rodent behavior is very important for achieving an in-depth understanding of behavioral and physiological phenomena, especially in nervous system, psychological disease, and drug development research. Observing and analyzing animal behavior helps researchers understand the functions and interactions of different neural pathways and study psychotherapeutic methods for treating diseases, mood disorders, and cognitive disorders.[1,2] The effectiveness and safety of drugs and how drugs affect the behavior and physiological responses of animals have been evaluated.[3–5] However, past animal behavior studies have often relied on experienced human observers, who must spend considerable time monitoring, evaluating, and quantifying animal behavior. Computer science and artificial intelligence (AI) algorithms, particularly deep learning technology, have been utilized in specialized experimental or home cage environments for research animals to comprehensively and accurately evaluate animal behavior and physiological status. AI has become increasingly prominent in animal behavior and neuroscience research. Researchers must observe experimental animals for extended periods in animal behavior studies, and video equipment is usually used to record the animals' behavioral responses.[6–8] When the experimental design requires long-term continuous observation, the researcher may become fatigued, leading to errors in judgment when evaluating complex or subtle animal behaviors. In addition, errors can occur due to differences among observers or laboratories, affecting the reliability and reproducibility of the research.[9–11] Moreover, collecting and analyzing large-scale data with manual methods is difficult when multiple animals must be observed simultaneously or recorded over long periods. Manual approaches are extremely time-consuming and inefficient, making it difficult to investigate complex behavioral patterns.

Notable advancements have been made in the development of animal behavior research techniques. Many automated monitoring methods, including radio frequency identification (RFID), infrared beams, capacitance technology, piezoelectric sensors, and video tracking algorithms, have been developed.[12] These methods have been applied to analyze the behavior and track the positions of rats and mice. Automated monitoring systems help researchers accurately capture and analyze these behaviors and enable long-term experiments. In addition, advances in automation help decrease researcher time and effort and enable large-scale research, thus accelerating scientific research. With the development of digital technology, automated monitoring systems with high-resolution photography and sensing technology have been used to continuously monitor animal behaviors.[13] Automated systems offer many benefits when used in animal behavior research. They are objective and consistent and can accurately capture and record a wide range of behaviors. These systems are becoming even more efficient and useful for providing detailed data for analysis with advances in AI.

The design of the experimental environment should be carefully considered to ensure that an animal's behavior is natural and reproducible. Researchers often house experimental animals in home cage environments, allowing the animals to exhibit more natural behaviors in a familiar, low-stress environment and reducing the distractions associated with laboratory environments. Over time, automated monitoring technology has been increasing applied in home cage environments.[14] The use of automated monitoring technology can greatly increase

[1]Laboratory Animal Center, National Health Research Institutes, 35, Keyan Road, Zhunan Town, Miaoli County 350401, Taiwan
[2]Department of Biomedical Engineering and Environmental Sciences, National Tsing Hua University, No. 101, Section 2, Kuang-Fu Road, Hsinchu City 300044, Taiwan
[3]Institute of Biomedical Engineering and Nanomedicine, National Health Research Institutes, 35, Keyan Road, Zhunan Town, Miaoli County 350401, Taiwan
[4]Lead contact
*Correspondence: jkchen@nhri.edu.tw (J.-K.C.), ldliao@nhri.edu.tw (L.-D.L.)
https://doi.org/10.1016/j.isci.2024.111223

the speed of analysis; moreover, this technology is objective and consistent, thereby ensuring the reproducibility and replicability of the research results.[15] Reproducibility and replicability are particularly important in behavioral research because animal behavior is often a key indicator. The use of automated monitoring techniques improves the standardization of the experimental design and reduces differences in the observation standards of different researchers or laboratories, which is crucial for ensuring reproducible research results. Existing automatic analysis methods focus mostly on observing rats and mice in home cage environments from a top or side view. A top-view camera design can achieve better results in recognizing the positions or interactive behaviors among multiple animals.[16–18] A side-view camera design can obtain clearer images of subtle movements and postures.[19,20] However, a top-view camera may not capture a complete view of the home cage environment because of obstruction from the cage cover or feed rack. A side-view camera needs to be installed outside the cage to obtain a complete view; however, image quality may be reduced in cages with low transparency, making it difficult to capture clear images of subtle behavioral postures. Therefore, in our design, the camera is installed in the cage to capture images from the front-view position, with the camera facing down at approximately 45° to monitor the rat. Setting the camera inside the cage is beneficial for obtaining high-quality images. This approach prevents the issue of the thermal lens being unable to detect body temperature due to the cage barrier. The front-view camera design enables the capture of subtle poses via the side-view camera and position tracking via the top-view camera.

Deep learning algorithms can be applied for the automatic recognition and analysis of animal behavior. Key point pose estimation methods have been used to measure animal movements in neuroscience research.[21–24] Pose estimation algorithms are used to automatically identify mouse body parts for animal behavior analysis. Deep learning algorithms can differentiate similar-looking objects to identify different species or behaviors. Liu et al.[25] used the neural network architecture of the SqueezeNet model to instantly identify *Aedes aegypti* and *Culex quinquefasciatus* in flight. Yu et al.[26] used an improved YOLOv3 model to detect the crawling behavior of ewes during estrus. Girardie et al.[27] used two YOLOv2 convolutional neural network (CNN) models to detect six postures and three standing activities by sows to evaluate how sow behavior affects piglet performance during early lactation. These results obtained using deep learning technology show that You Only Look Once (YOLO)-based object detection algorithms can be used to quantitatively analyze complex behaviors and track animal positions. In this study, we used the YOLOv5[28] deep learning network architecture to build the model. YOLO algorithms are famous for their excellent real-time recognition capabilities and efficient processing speed,[29] which are crucial for large-scale animal behavior monitoring and analysis. YOLOv5 also performs well in terms of object position detection and classification accuracy. The advantages of YOLOv5 include the ability to develop behavioral recognition models for rats. Objects and their positions can be identified instantly during a single trial with high accuracy, thus improving research efficiency and the objectivity of the results compared with manual work.

Rodents remain the primary experimental animals utilized during the preclinical research phase of drug development.[30,31] The European Parliament implemented strengthened animal protection policies through Directive 2010/63/EU, mentioning the principles of replacement, reduction, and refinement (3Rs),[32] and 3R-related platforms continue to be established.[33] The 3R principle is a universal concept among international researchers. Routine disinfection and cleaning of experimental animal materials can be mechanized to reduce human workload. However, daily clinical observation and care of animals must still be performed via human monitoring. Timely intervention, treatment, and setting of humane endpoints on the basis of an animal's health status require considerable human workload and time. In experimental animal facilities with large animal populations, caretakers and researchers are limited in the number of daily observations of experimental animals they can perform, making it challenging to monitor the health of each animal continuously. Moreover, during the outbreak of large-scale infectious diseases, medical resources can be scarce, and home care and isolation are prevalent. Remote real-time physiological monitoring devices can be used to help monitor patients with mild symptoms or asymptomatic patients, providing early warning capabilities and timely medical guidance.[34–36] Furthermore, these devices can detect both physical parameters and psychological stress to assess emotional and stress levels, allowing intervention before the condition worsens.[37–39] Recent technological advances and a focus on animal welfare have led to the development of real-time physiological monitoring systems for laboratory animal facilities. This study presents the RodentWatch (RW) system, which is small, scalable, and suitable for standard cages and can detect behavior, location, and body temperature data. Future updates will include the development of abnormal behavior recognition and early warning systems in accordance with the 3R principles to ensure animal welfare.

In this study, the RW system was developed, and the YOLOv5s neural network architecture was used to establish a real-time behavior recognition model for rats. The hardware is lightweight and easy to install, with a single multifunction camera design. The compact design allows the camera to be installed in the cage, and long-term monitoring of the animals in the original cage can be performed during both day and night without moving the animals. The designed hardware system is simple and inexpensive, which can reduce commercialization costs and increase the potential for expansion. We expect this tool to be used in preclinical research and daily care in laboratory animal facilities.

RESULTS

We developed the RW system using a multifunctional camera and an edge computing device to identify rat behaviors through deep learning models. Figure 1 illustrates the hardware design of the RW system, while Figure 2 presents its operational and processing framework. Figure 3 provides representative examples of the eight behavioral posture categories. Figure 4 offers an overview of the dataset, which includes training, validation, and test sets. Our study focused on identifying drinking and resting behaviors in rats, with recognition accuracy enhanced through contextual object labeling and modified non-maximum suppression (NMS) strategies. Two experiments were conducted to evaluate the usability of this system in rat behavioral studies. These aspects are described in detail in the following sections and the STAR Methods.

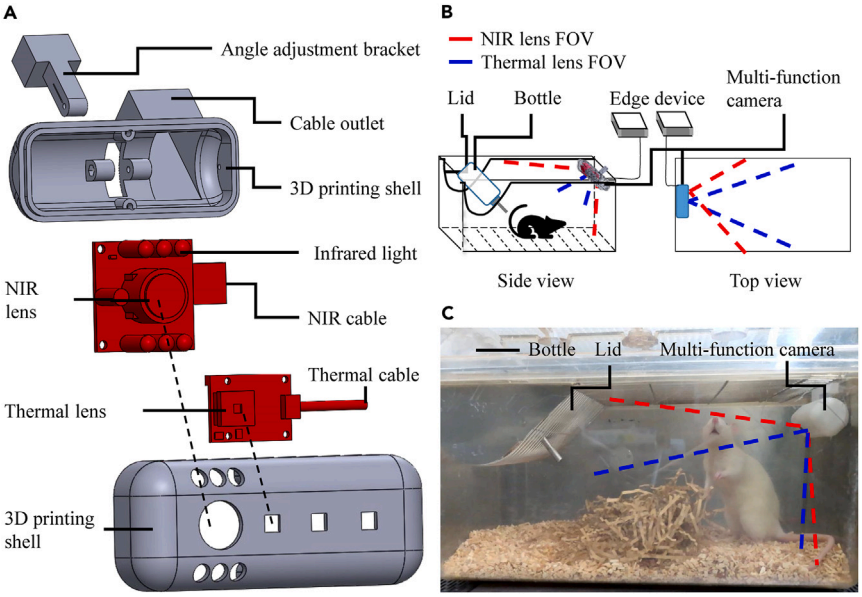


Figure 1. Overview of the RW system

(A) Exploded-view schematic of the single multifunctional camera. The multifunctional camera comprises a near-infrared (NIR) lens, near-infrared light, a thermal lens, a 3D printing shell, and an angle adjustment bracket.

(B) Schematic of the multifunctional camera mounted in a cage. The camera monitors the rat at an angle of approximately 45° and is connected to the edge computing device through a wire. The NIR lens FOV is greater than 160° (red lines). The thermal lens FOV is 57° (blue lines).

(C) Actual image of the camera mounted in the cage. The camera is simple and convenient to mount in a standard housing cage for rats. The modified cage lid reduces the problem of the feeder position blocking the view of the rat.

Enhanced recognition of drinking behavior

Establishing the YOLOv5s model requires manual annotation of bounding boxes and behavior categories to create datasets. The manually annotated bounding box should be as close as possible to the target object to prevent unnecessary background information from affecting the feature extraction process of the model. We utilized the pretrained YOLOv5s deep learning network architecture to construct the original model (O model). We utilized the contextual object labeling method with the training set to expand the object bounding boxes to include drinking behavior and the bottle cap area to improve the accuracy of the model in recognizing drinking behavior. The drinking category in the training set was reannotated, and an enhanced drinking model (D model) was developed using the same parameter settings. We used the evaluation tools provided by YOLOv5 to assess the performance of the two models in the object detection task on the basis of the validation sets.

The AP metric is suitable for evaluating performance differences between models. The AP values of each category under the same IoU threshold of 0.5 were compared between the O and D models. The results show that the AP value of the drinking category increased from 0.785 for the O model to 0.909 for the D model, as shown in Figure 5A. The AP value of the resting category also increased from 0.869 for the O model to 0.911 for the D model, as shown in Figure 5H. In addition, the AP values of the grooming, searching, and gnawing categories increased to varying degrees from the O model to the D model, as shown in Figures 5C, 5E, and 5F. The AP values of the eating and rearing categories decreased slightly from the O model to the D model, as shown in Figures 5B and 5D. Compared to the O model, the accuracy of identifying the scratching category improved with the D model, with an AP value below 0.1, as shown in Figure 5G. The average AP values of all categories (mAP) improved from 0.722 for the O model to 0.771 for the D model, as shown in Figure 5I. The results show that expanding the bounding box representing the drinking behavior to include contextual objects can enhance model accuracy.

A confusion matrix is a tool used to evaluate the prediction performance of machine learning classification models. Analyzing the correspondence between the model's predictions based on the validation set and the actual categories is beneficial for understanding the model's performance in each behavioral category. In a confusion matrix, the rows represent the actual behavior, whereas the columns represent the behavior predicted by the model. In a binary classification problem, the values in the confusion matrix denote the proportions of instances from each true class that are classified into the predicted categories. Specifically, in Figure 6A, the value of 0.67 for the true category "drinking" predicted as "drinking" indicates that 67% of the instances from the true "drinking" category were accurately classified as TPs by the model. Thus, each cell in the confusion matrix represents the proportion of instances from a given true class that are assigned to a specific predicted class. The O model correctly predicted the numbers of positive categories, including eating, rearing, and searching, with a detection rate exceeding 0.8, whereas the detection rates for other categories were below 0.7. The detection error rate for the resting category being incorrectly predicted as the searching category was as high as 0.91 since the behavioral features of the two categories are very similar,

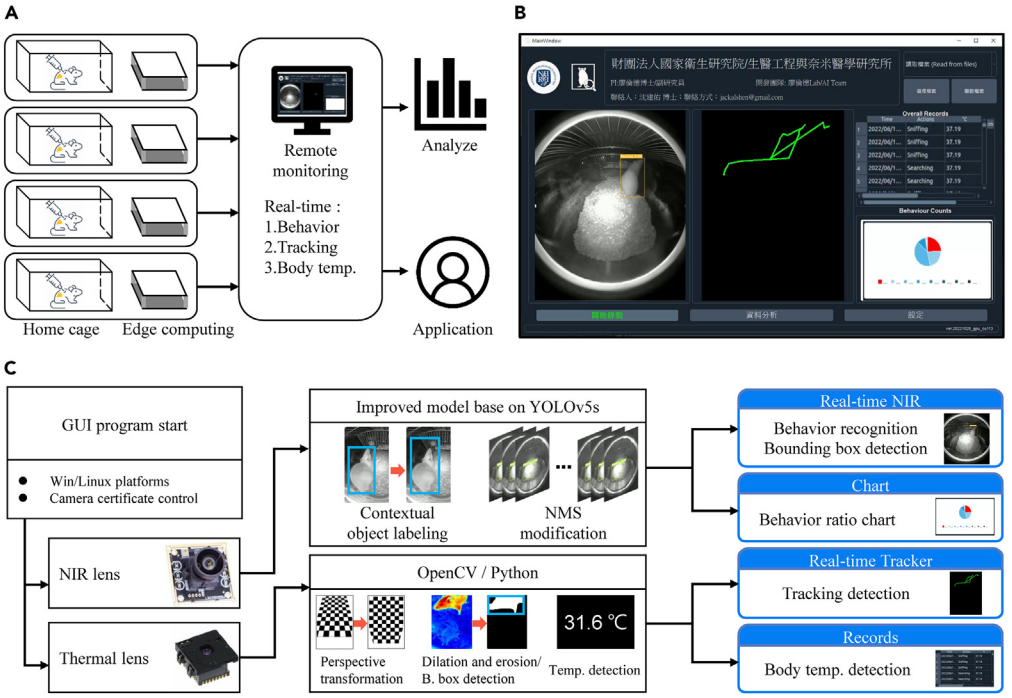


Figure 2. The operation and processing framework of the RW system

(A) Operation and application of the RW system. The hardware is lightweight and can be easily expanded to multiple housing cages. The housing cage can be monitored on the original cage rack without moving the cage. The burden of manual observation on researchers is reduced through visual real-time detection with the remote monitoring system. Complete video and data recording allow for further analysis and related applications.

(B) Graphical user interface (GUI) software tool for the RW system. We developed this tool using Python and PyQt5 to provide a convenient user interface for easy operation, real-time information display, and data recording.

(C) Flowchart of the proposed method. The NIR lens is used to acquire rat images, and the improved YOLOv5 model is used to identify behaviors and bounding boxes. The thermal lens is used to track position and body temperature changes. Real-time detection results can be displayed via the GUI.

as shown in Figure 6A. Compared with the O model, the enhanced D model has a higher correct detection rate, particularly for the drinking category, with the detection rate of this category increasing from 0.67 to 0.90. In addition, the detection rate of the resting category was higher for the D model, increasing from 0.08 to 0.48, as shown in Figure 6B. Since the training set includes many images in the searching category, both models tend to predict the searching category. This results in a lower recall rate for the gnawing and scratching categories because there are fewer images representing these behaviors in the training set. However, the improved D model still achieved a correct detection rate of more than 0.85 for four behavioral categories: drinking, eating, rearing, and searching.

The drinking and resting categories can easily be differentiated by the model on the basis of the identified instances in the validation set images. Multiple detection bounding boxes are generated in the images with the confidence threshold set to 0.001, allowing us to intuitively understand the recognition results of the model. Figure 6C shows an image that represents the true drinking behavior category identified by the O model. The model predicted three categories with confidence values of 0.86, 0.004, and 0.002. However, the drinking category was not successfully identified by the model, and resting behavior was incorrectly identified. The recognition results of the D model for the image are shown in Figure 6D. The model predicted two categories: drinking, with a confidence value of 0.72, and searching, with a confidence value of 0.42. The results show that the improved D model can better detect the drinking category, and the incorrect detection of the resting category is reduced. In addition, Figure 6E shows an image that represents the true resting behavior identified by the O model. The model predicted two categories with confidence values of 0.64 and 0.002. However, the resting category was not successfully identified, and the drinking category was incorrectly identified. The recognition results of the D model for the image are shown in Figure 6F. The model predicted three categories with confidence values of 0.75, 0.17 and 0.03. The results show that the improved D model can better detect the drinking and resting categories, and the incorrect detection of the drinking category is reduced.

Enhanced recognition of resting behavior

The F1-confidence curve is used to evaluate the performance of binary classifiers, considering the classification accuracy and recall, as well as the impact of different confidence thresholds. The performance of the D model was improved by incorporating the situational object labeling method, which was used for detection based on the verification set. The F1 scores of each category under different confidence thresholds are

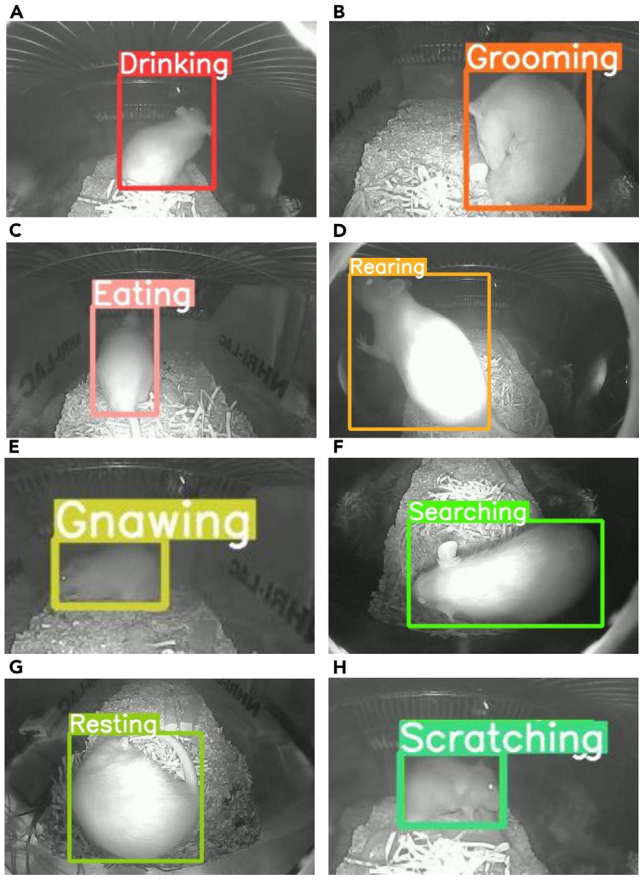


Figure 3. Representatives of 8 behavioral posture categories

The behavioral categories in this study included (A) drinking, (B) grooming, (C) eating, (D) rearing, (E) gnawing, (F) searching, (G) resting, and (H) scratching. Searching is defined as the default behavior and includes unclassified behaviors and partial images with limited viewing angles.

shown in Figure 7A. The model showed the best performance for the behavior category, with the average highest F1 score of 0.64, corresponding to a confidence threshold of 0.395. This result demonstrates that the model achieves a good balance between precision and recall at this confidence threshold, making it suitable for application in actual scenarios. Therefore, the NMS modification, test set, and animal experiment results were analyzed on the basis of these confidence values.

The stationary features of the rats were used to improve the model's accuracy for the resting category. First, a set of fewer than 300 detected instances was output by the D model. For these instances, the confidence values for the resting and searching categories were adjusted on the basis of whether the rat was stationary. The detected instance with the highest confidence score after NMS processing was subsequently used as the final output. The D model with the modified NMS strategy was named the resting and drinking enhanced model (RD model). The validation set results revealed that the F1 score of the resting category increased from 0.570 with the D model to 0.678 with the RD model and that the F1 score of the searching category increased from 0.742 with the D model to 0.766 with the RD model. The average F1 score of all the behavioral categories also increased from 0.644 to 0.664, as shown in Figure 7B.

The images in the validation set were examined to understand the impacts of the two models on detecting resting and searching behaviors. Figure 7C shows six consecutive images of a rat displaying actual resting behavior, with 1 s intervals between images. The D model correctly detected resting behavior in only one image, with a confidence threshold of 0.395, and the other images were misidentified as containing searching behavior. In contrast, all six images were correctly identified as showing resting behavior by the RD model. The results for this representative image show that the RD model can increase the confidence of the resting category and reduce the confidence of the searching category when the rat is in a stationary state, thereby increasing the correct detection rate.

In addition, upon further analysis with the confidence threshold set at 0.395 for the O, D, and RD models, the overall F1 scores of the three models are 0.566, 0.644, and 0.664, respectively. The overall recognition accuracy is effectively enhanced with these model improvements. We conducted McNemar's test to evaluate the differences in the recognition accuracy among the models. The results demonstrated a significant difference in accuracy between the O and D models ($p = 4.36 \times 10^{-130}$, McNemar's test), as well as between the D and RD models ($p = 8.75 \times 10^{-68}$, McNemar's test).

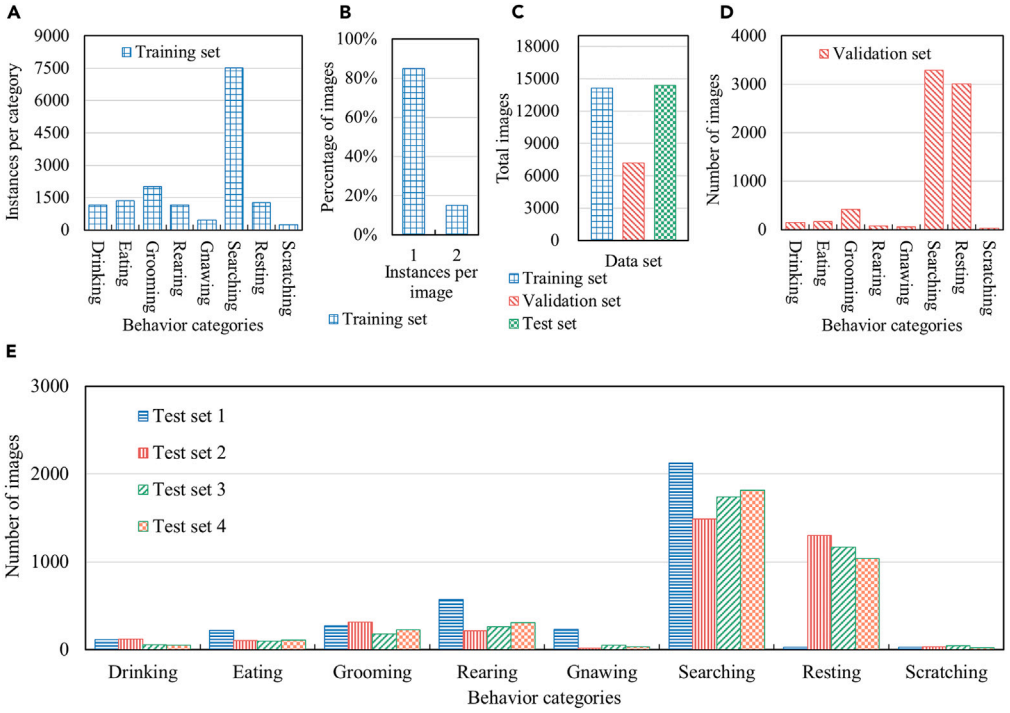


Figure 4. Dataset overview

(A) The number of instances of each category in the training set. Under normal conditions, the dominant behavioral activity is searching, and many instances of this category are included in the training set. In addition, fewer gnawing and scratching behaviors are included.

(B) The percentage of instances in each image in the training set. The single-rat images in the training set accounted for 84.9% of the total images, and the two-rat images accounted for 15.1% of the images.

(C) The number of images in the training, validation, and test sets. The training set includes manually selected images from multiple videos. The images in the validation and test sets include continuously captured images taken at 1-s intervals.

(D) The number of images for each category in the validation set. A high proportion of the images in the validation set were in the searching and resting behavior categories.

(E) The number of images for each category in the test set. The test set includes test sets 1, 2, 3, and 4. The test sets included more images of searching and resting behaviors than images of other behaviors.

Performance analysis based on the test set

The main objective of including the test set is to test the model's performance with unseen data to determine whether the model can generalize well to real-world data, as opposed to the validation set. The test set includes test sets 1, 2, 3 and 4, and the behaviors in the images were detected by the RD model. The results revealed that the F1 scores of the drinking, eating, rearing, searching, and resting categories were greater than 0.8, and the F1 scores of the drinking, eating, and rearing categories were greater than 0.9. The average F1 score of all categories was 0.756, as shown in Figure 8A. The confusion matrix analysis results were similar to those of the F1 score. The correct detection rate by the RD model was greater than 0.8 for the drinking, eating, rearing, searching, and resting categories, with the correct detection rate for the drinking, rearing, and resting categories exceeding 0.9, as shown in Figure 8B.

A detailed analysis of the precision and recall for each category based on test sets 1, 2, 3, and 4 is shown in Table 1. The precision in identifying grooming behavior was greater than 0.8 with all four test sets; however, the recall was relatively low with test sets 1, 2, and 4. The F1 scores for each category in test sets 1, 2, 3, and 4 are shown in Table 2. Specifically, the F1 scores of the grooming category were 0.793, 0.432, 0.894, and 0.676 for test sets 1, 2, 3, and 4, respectively. The disparities in the F1 values among the test sets may be attributed to the diversity of grooming behaviors or individual differences in the rats. Expanding the dataset could enhance identification performance. The F1 scores of the resting category were 0.409, 0.851, 0.960, and 0.812 for test sets 1, 2, 3, and 4, respectively. This inconsistency may be due to differences in the number of images in the test sets. The low accuracy in identifying the scratching category may be attributed to the insufficient number of images in the training, validation, and test sets. Overall, the RD model had an F1 score higher than 0.8 in identifying five categories in the test set, and the confusion matrix results demonstrated the high correct detection rate of the RD model. These results show that the contextual object labeling and enhanced NMS methods effectively improve the model's accuracy in identifying drinking and resting behaviors, as well as the overall detection performance of the model. Thus, this model may have good generalizability in practical applications.

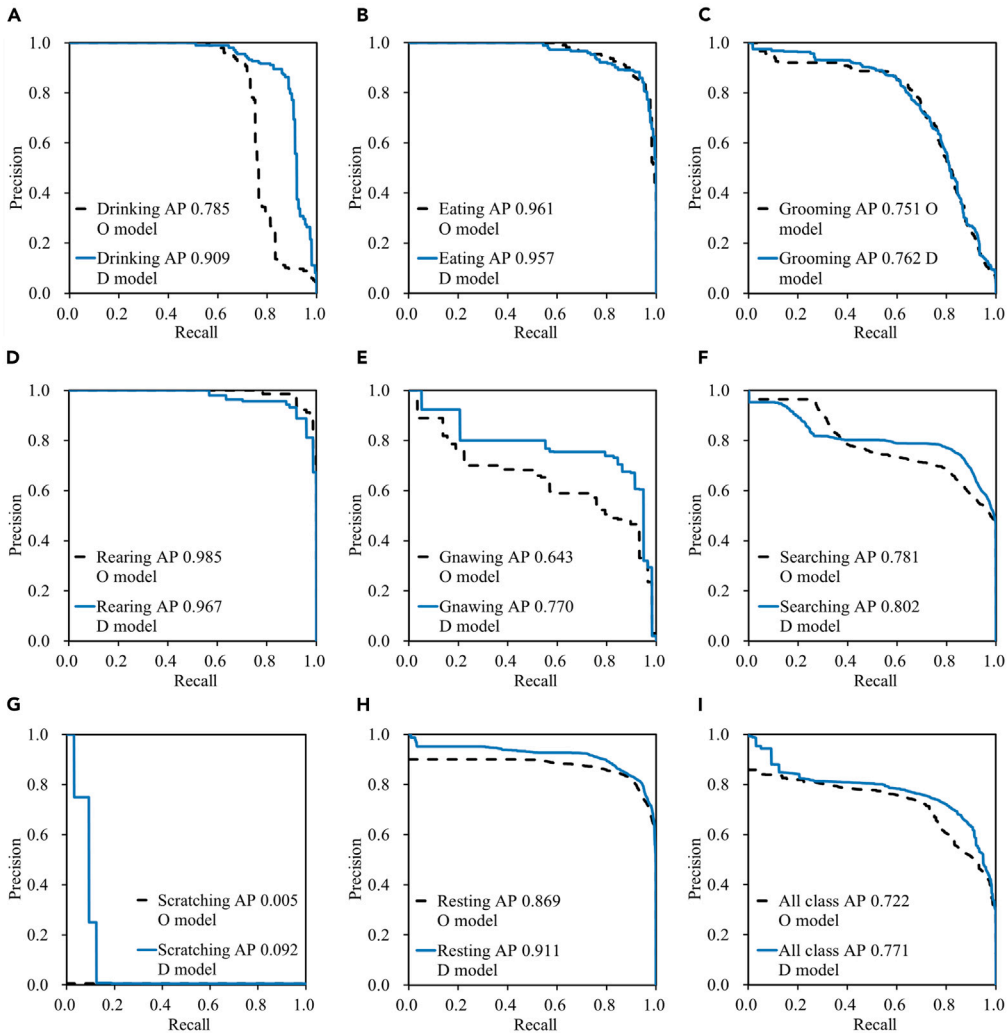


Figure 5. Comparison of the PR curves of the O and D models

(A–H) PR curves for each of the 8 behavioral categories. The AP values of the O and D models were compared, with the IoU threshold set to 0.5. The AP value of the drinking category increased from 0.785 for the O model to 0.909 for the D model. The AP value of the resting category increased from 0.869 for the O model to 0.911 for the D model. In addition, the AP values of the grooming, searching, and gnawing categories increased to varying degrees. However, the AP values for the eating and rearing categories decreased slightly, and the precision of the scratching category score was insufficient.

(I) The average AP value of all categories with the IoU threshold set to 0.5 increased from 0.722 for the O model to 0.771 for the D model.

Application of the developed RW system

The RW system developed in this study can be used to record behavior, position, and body temperature data of rats to further analyze the rat's behavioral patterns and body temperature changes. For example, in the analysis based on test set 1, we generated a fence diagram according to each category of rat behavior every second, as shown in Figure 9A (See also Video S1). A stacked bar chart can also represent the proportion of each behavior at 5-min intervals, as shown in Figure 9C. The time of occurrence and duration of each behavior can be fully represented through such plots. Furthermore, heatmaps are typically used to determine in which specific areas rats tend to linger for extended periods. This helps in identifying the rats' preferred locations or specific activity areas, making it easier for researchers to comprehend the rats' behavioral patterns and spatial distribution, as shown in Figure 9B. In addition, the body temperature of each rat was monitored and recorded using a thermal lens. The temperature records in test set 1 were processed with a moving average, and the change curve is shown in Figure 9D. A body temperature change chart can be used to study the biological behavior of rats and explore the relationship between physiology and body temperature changes. In physiological research, these graphs help analyze how internal and external factors affect a rat's body temperature.

Experiment 1: Light cycle change experiment

The L/D group had a normal light cycle of 12 h light/12 h dark on the first day of the experiment. The results revealed that the major behavior exhibited by the animals during the day was resting, and the activity frequency was low, with the most active time being night, as shown in

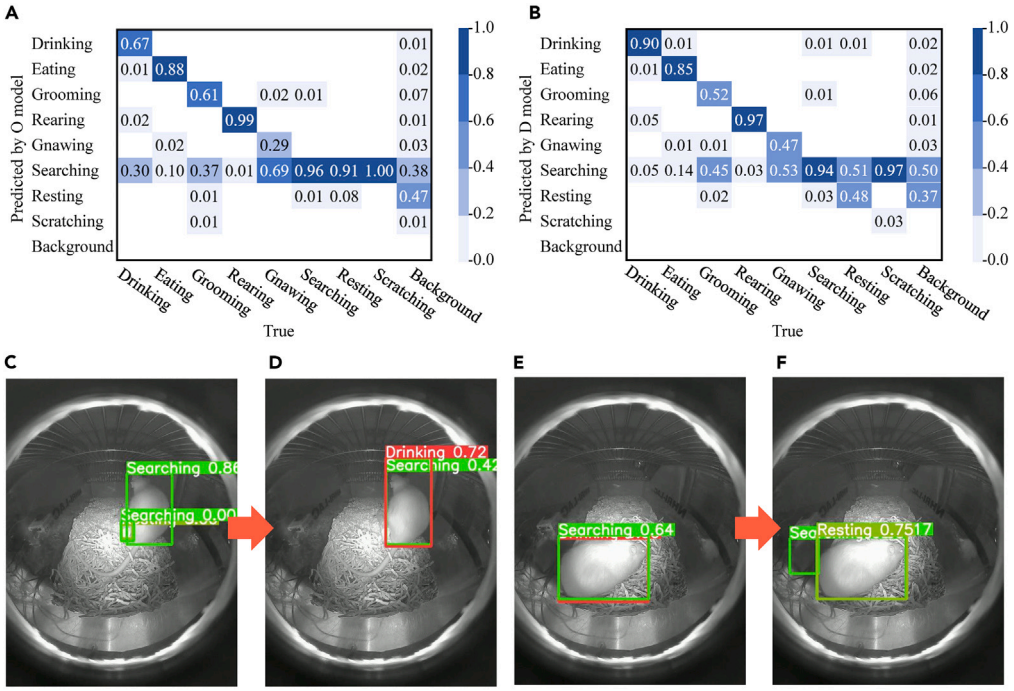


Figure 6. Comparison of the confusion matrix and behavior recognition results of the O model and the D model

(A) Confusion matrix of the O model. The detection rates of correctly predicted positive categories for the eating, rearing, and searching categories were greater than 0.8. Since the behavioral posture features of these behaviors are very similar, the rate of incorrectly predicting the true resting category as the searching category was as high as 0.91.

(B) Confusion matrix of the D model. The correct detection rates of the improved D model for the drinking, eating, rearing, and searching categories were greater than 0.85. The correct detection rate of the drinking category increased significantly from 0.67 to 0.90, which also caused the correct detection rate of the resting category to increase significantly from 0.08 to 0.48.

(C) Representative image of the drinking category identified by the O model with the confidence threshold set to 0.001. The O model identified three objects but did not correctly predict the drinking category.

(D) The same image as (C) was correctly identified by the D model. The D model identified two objects, one of which was correctly predicted as the drinking category.

(E) Representative image of the resting category identified by the O model with the confidence threshold set to 0.001. The O model identified two objects but did not correctly predict the resting category.

(F) The same image as (E) was identified by the D model. The D model identified three objects, one of which was correctly predicted as the resting category.

Figure 10A. In the L/L group, the light cycle was 12 h light/12 h light on the second day of the experiment. Compared with that in the L/D group, the activity frequency of the rats in this group decreased at night, indicating that light affects rat activity, as shown in Figure 10B. We analyzed the differences in daytime and nighttime rat activity. The results revealed that during the daytime, the rats in both the L/L and L/D groups were active less than 3.5 $\pm$ 0.4% of the time regardless of the light cycle conditions. However, at night, the rats in the L/D group were active 18.5 $\pm$ 1.8% of the time, which was significantly greater than the value of 6.8 $\pm$ 1.1% in the L/L group ($p = 9.90 \times 10^{-5}$, *t* test, $n = 7$), as shown in Figure 10C. The results indicate that the behavioral distributions of the rats in the L/D and L/L groups were consistent with the nocturnal habits of rodents. In addition, the natural habit of rodents is to rest or have low activity levels during the day. The analysis revealed that resting behavior accounted for 61.8 $\pm$ 7.8–66.7 $\pm$ 2.9% of the behavior during the day. In contrast, in the L/D group, for which the lights were turned off for 12 h at night, resting behavior accounted for only 32.7 $\pm$ 4.7% of the overall behavior, as shown in Figure 10D. However, when the light was turned on at night, the resting behavior of the rats in the L/L group significantly increased to 49.1 $\pm$ 3.0% of the overall behavior ($p = 0.001$, *t* test, $n = 7$).

The thermal lens location data of the rats can be used to generate a heatmap of the locations of the rats in the home cage. The housing cages were placed on a shelf against a wall in the housing room. We divided the housing cage into two areas to analyze the response of the rats to brightness. Zone A represents the side closer to the wall, where the laminate blocks the indoor light, and the brightness is low. Additionally, zone B represents the area near the outside of the cage, which is brighter. The L/D group and L/L group were subjected to different light cycle conditions on the first and second days, respectively. The heatmap shows that the activity levels of the rats in both groups were lower during the day, and the preferred location was zone A of the housing cage, as shown in Figures 10G1 and 10G3. The rats in the L/D group were active in zones A and B when the light was turned off at night, as shown in Figure 10G2. However, the rats in the L/L group were affected by light when the light was turned on at night, and the preferred location was zone A of the housing cage, as shown in Figure 10G4.

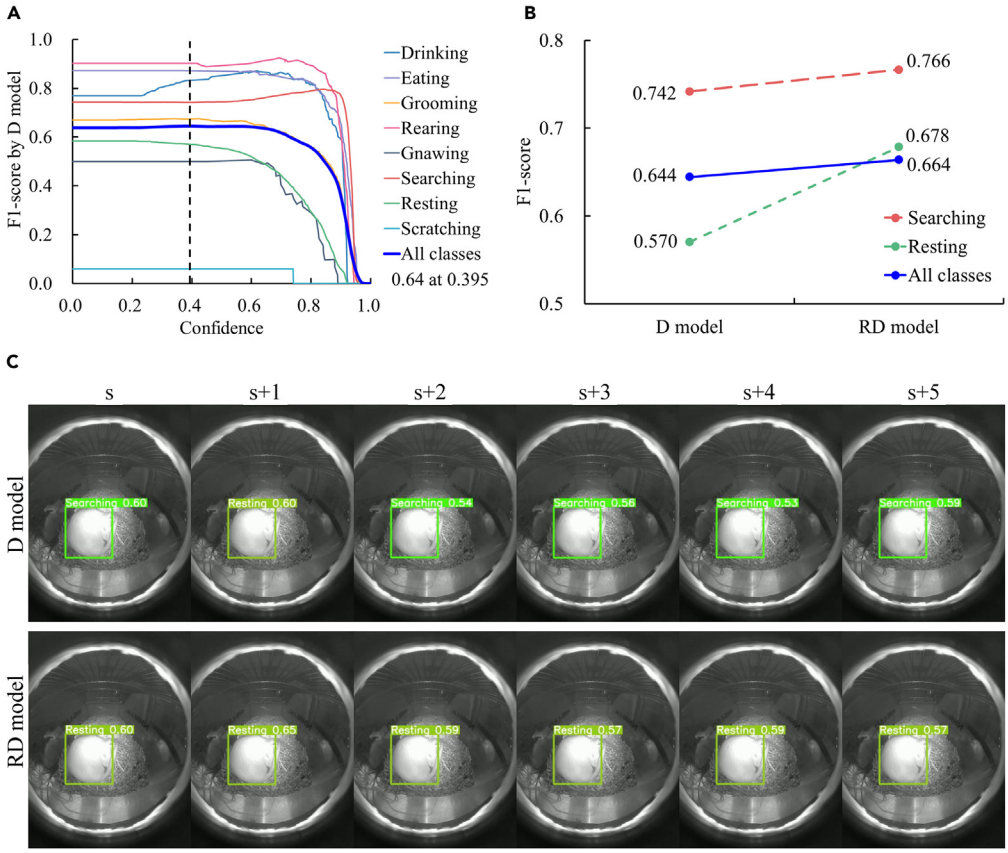


Figure 7. Comparison of the F1 score and behavior recognition results of the D model and RD model

(A) F1-confidence curve of the D model based on the validation set. The behavioral category with the highest average F1 score of 0.64 corresponds to a confidence level of 0.395. This indicates that the model achieves a good balance between precision and recall at this confidence threshold.

(B) F1 scores of the D and RD models with the confidence level set to 0.395. The validation set results revealed that the F1 score of the resting category increased from 0.570 with the D model to 0.678 with the RD model and that that of the searching category increased from 0.742 with the D model to 0.766 with the RD model. The average F1 score of all the behavioral categories also increased from 0.644 to 0.664.

(C) Representative image of the resting category identified by the D and RD models with the confidence set to 0.395. Six consecutive representative images of the resting category were captured continuously at one image per second. The D model correctly identified only one image, whereas the RD model correctly identified all six images as the resting category.

A detailed analysis was conducted to compare the location distribution differences between the rats in zones A and B when the lights were turned off at night. The distribution of the locations of the rats in the L/D group in zones A and B were similar, indicating that the rats were active and moved around the housing cage when the lights were off at night. However, when the lights were turned on at night, the rats in the L/L group preferred zone A, which had less light, compared with those in the L/D group ($p = 4.65 \times 10^{-10}$, *t* test, $n = 7$), as shown in Figure 10E. In addition, the thermal lens recorded the changes in the body surface temperature of the rats during the day and night. The average body surface temperatures of the rats in the L/D group during the day and night were 30.8 $\pm$ 0.32°C and 33.1 $\pm$ 0.22°C, respectively. In contrast, the average body surface temperature of the rats in the L/L group was 31.1 $\pm$ 0.16°C during the day and 31.1 $\pm$ 0.33 °C at night. These results indicate that under the standard light cycle, the body temperatures of the rats in the L/D group were greater at night than during the day ($p = 7.46 \times 10^{-5}$, *t* test, $n = 7$). However, when the lights were on at night, there was no significant difference in the body temperature of the rats in the L/L group at night and during the day ($p = 0.93$, *t* test, $n = 7$), as shown in Figure 10F.

Overall, the results showed that in the L/D group under a standard light cycle, the rats exhibited typical nocturnal activity patterns and rested more during the day. However, the activity levels of the rats in the L/L group with the lights turned on at night were significantly lower than those of the rats in the L/D group, resulting in no significant difference in the preferred positions of the rats in the L/L group during the day or at night. In addition, the body temperatures of the rats in the L/L group were lower than those of the rats in the L/D group when the lights were turned on at night, suggesting that changes in the light cycle impacted the body temperatures of the rats.

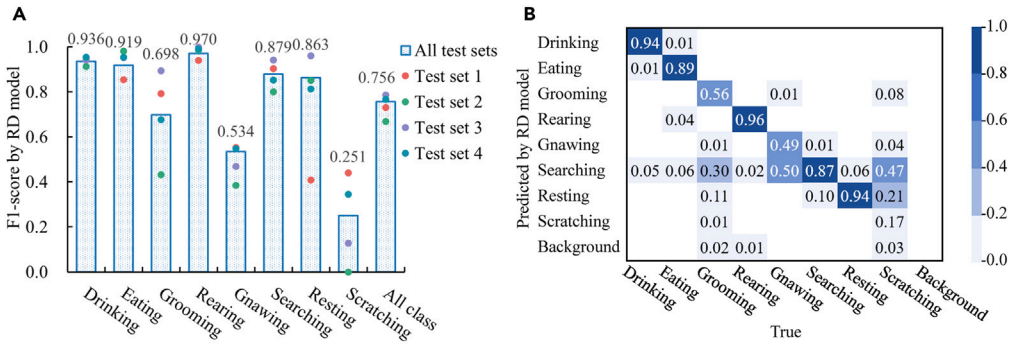


Figure 8. Evaluation of the RD model based on the test set

(A) The F1 scores of the RD model were evaluated across the four subsets of the test set. The F1 scores of the drinking, eating, rearing, searching, and resting categories were all greater than 0.8 based on all test sets. The average F1 score across all categories was 0.756.

(B) Confusion matrix of the RD model. The confusion matrix results show five behavioral categories with correct detection ratios higher than 0.8.

Experiment 2: Sucrose preference test

For the SPT, we analyzed and compared the results of three groups, namely, the W/W group, S/S group and W/S group. The experimental process is shown in Figure 11A. The results showed that the rats in the W/W group spent 409 $\pm$ 106 and 644 $\pm$ 126 s drinking water from the two sucrose-free water bottles, with no significant difference between groups ($p = 0.48$, *t* test, $n = 8$). On the second day, the amount of water consumed by the rats from the two bottles of sucrose-containing water was significantly greater in the S/S group than in the W/W group (2,055 $\pm$ 347 and 2,374 $\pm$ 406 s, respectively). On the third day, the rats in the W/S group spent 2,992 $\pm$ 666 s drinking from the sucrose-containing water bottle on the right side, which was significantly longer than the 166 $\pm$ 37 s spent drinking from the sucrose-free water bottle on the left side ($p = 3.56 \times 10^{-3}$, *t* test, $n = 8$), as shown in Figure 11B. The sucrose preference of the rats in the W/S group was 94.3 $\pm$ 1.3%, indicating that the rats clearly preferred sucrose, as shown in Figure 11C.

Some rats may have a position preference in the SPT. The effect of this position preference can be eliminated by changing the position of the water bottle.[40] The position of the water bottle was not changed during this experiment. The data obtained from the eight rats in the W/W group and S/S group were analyzed individually, which revealed that some rats had different degrees of water bottle preference, as shown in Figures 11D and 11E. However, the *t* test results of the W/W ($p = 0.48$, *t* test, $n = 8$) and S/S ($p = 0.14$, *t* test, $n = 8$) groups revealed no significant difference in the time spent drinking from the left or right water bottle, as shown in Figure 11B. The purpose of the W/W and S/S groups in the two experiments was to teach the rats the difference between water bottles with and without sucrose. Then, an SPT was conducted with the W/S group. The results of this experiment revealed that the eight individual rats had clear preferences for sucrose in the right water bottle containing sucrose, and the average sucrose preference was 94.3 $\pm$ 1.3%, as shown in Figure 11F.

DISCUSSION

Enhancing behavior recognition through contextual object labeling

The bounding boxes in the object detection task must be manually annotated. The bounding box usually surrounds the object as closely as possible to promote model generalization, which allows the model to learn the object features and reduce interference from background information. However, the drinking behavior of the rats was misidentified as searching behavior. The inadequate accuracy in identifying drinking behavior may be attributed to the model learning insufficient features. In actual images of rats showing drinking behavior, the image always includes a water bottle cap. Therefore, the contextual object labeling method was used to expand the bounding boxes for drinking behavior to include rats and water bottle caps. Thus, the model learns the contextual features of water bottle caps, which helps enhance its ability to identify drinking behaviors.

In addition, the improved model could more accurately identify the resting category. The AP value of the resting category for the O model, which does not include the water bottle cap feature, was 0.869 based on the validation set. In contrast, the AP value of the resting category for the D model, which does incorporate the water bottle cap feature, increased to 0.911, as shown in Figure 5H. For the resting category, the O model had a high precision value and a low correct detection rate of only 0.08, as shown in Figure 6A. The low correct detection rate was caused by the low recall, meaning that the O model identified too many FNs in recognizing resting behavior. In the D model, the recall value for identifying resting behavior was improved, and the correct detection rate increased to 0.48. This finding indicates that the D model can more accurately identify both drinking behavior and resting behavior, and the number of incorrect searching behavior identifications is reduced, as shown in Figure 6B. In the SPT, accurate identification of drinking behavior by the model is critical. The identified behavior and location information revealed which bottle the rat drank from and at what time. The sucrose preference of the rats was consistent with the results of the typical SPT. Therefore, incorporating the bottle cap as a situational object into a rat's bounding box is a reasonable approach for improving the accuracy of the model in identifying a rat's drinking behavior. The results of this experiment verified that the model is practical and can simulate human object recognition in the real world.

Table 1. Analysis of precision and recall of the RD model based on the test set

	Test set 1		Test set 2		Test set 3		Test set 4		All test sets	
Class	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Drinking	0.920	0.983	0.948	0.879	0.918	0.966	0.927	0.981	0.930	0.943
Eating	0.913	0.803	0.971	0.990	0.978	0.929	0.963	0.945	0.948	0.891
Grooming	0.918	0.697	0.899	0.284	0.974	0.826	0.959	0.518	0.935	0.557
Rearing	0.945	0.934	0.995	0.977	1.000	0.996	0.997	0.982	0.976	0.965
Gnawing	0.651	0.481	0.556	0.294	0.524	0.423	0.433	0.743	0.592	0.487
Searching	0.894	0.912	0.863	0.746	0.936	0.947	0.838	0.869	0.883	0.874
Resting	0.257	1.000	0.740	1.000	0.937	0.983	0.809	0.817	0.799	0.938
Scratching	0.419	0.464	0.000	0.000	1.000	0.068	0.542	0.263	0.500	0.168
All class	0.740	0.784	0.747	0.646	0.908	0.767	0.808	0.765	0.821	0.728

The precision and recall values of the drinking, eating, and rearing categories were consistently high across all test sets, with the precision values ranging from 0.913 to 1.000 and recall values ranging from 0.803 to 0.996. In contrast, the precision and recall of the gnawing and scratching categories were lower, with mean values ranging from 0.168 to 0.592 across all test sets, indicating challenges in accurately classifying these behaviors. The aggregated results for all categories revealed average precisions ranging from 0.740 to 0.908 across individual test sets, with an overall average precision of 0.821 across all test sets.

Enhancing behavior recognition with a modified NMS strategy

In the CNN layer, the input image is processed to extract features and obtain visual information about the different edges, textures, and positions of an object on the basis of its appearance and characteristics. The feature information extracted from images showing highly similar rat behavioral postures was insufficient as training data, and the model could not achieve good recognition accuracy. According to the behavioral classification of the rats in this study, the postures of resting and searching behaviors were very similar. The modified NMS strategy considers the stationary states of objects as process feature and adjusts the confidence scores of the resting and searching categories to improve accuracy. In addition to resting behavior, various rat behaviors involve static postures that are continuous across frames. We set the optimal parameters and thresholds on the basis of the detection results obtained based on the verification set to achieve the best accuracy and thus improve the recognition performance of the model. The accuracy and generalizability of the modified RD model were confirmed based on the test set, as shown in Figure 7B. We also observed that the F1 score of the resting category based on test set 1 was lower than the F1 scores of this category based on the other three test sets, which may be due to the different numbers of samples in each test set. The F1 score of the resting category based on all test sets was 0.863, which was higher than the 0.678 result based on the validation set. This may be related to individual differences among rats, and increasing the number of samples in the dataset may improve the robustness of the model.

The model's accuracy in identifying resting behavior was a critical indicator in the light cycle change experiment, given the close correlation between resting behavior and lighting conditions. The experimental results showed that the model successfully identified resting behaviors with or without light at night and fully reflected the natural habits of the rats. Therefore, specific optimization processes may be required in object detection tasks in certain scenarios to ensure better detection of target objects. The modified NMS strategy considers the characteristics of a rat's stationary state to optimize model performance. This method was useful for improving the recognition accuracy of rat behavior. These results demonstrate the value of our approach in practical applications, indicating that our model offers a promising solution for similar behavior recognition tasks.

Consistency of experimental validation

We conducted two animal experiments to verify the usability of the RW system in practical applications. In related research on light cycle changes, Depres-Brummer et al.[41] reported that long-term light cycles can induce complete suppression of body temperature and activity rhythm changes. Our experimental results are consistent with these findings, showing that light reduces the active behavior and body surface temperature of rats and increases their resting behavior, as shown in Figures 10C and 10F. In addition, in the L/D group with a standard light cycle, resting behaviors during the day accounted for 61.8–66.7% of all behaviors. In contrast, resting behaviors accounted for only 32.7% of all behaviors at night. This result is consistent with the findings of Ikeda et al.,[42] who reported that the total percentage of non-rapid eye moment (NREM) and rapid eye movement (REM) sleep during the daytime was 68%, whereas the percentage at night was 35.0%.

Relevant studies have shown that the body temperatures of rats at night are greater than those during the day[43–45] and that nighttime light stimulation can decrease body temperature.[46] This finding is consistent with our experimental results, which showed that the rats' body temperature increased when the light was turned off at night. In contrast, the rats' body temperatures were reduced by light stimulation when the light was turned on at night, as shown in Figure 10F. However, the surface body temperatures of different body parts of rats vary, with higher surface temperatures around the eyes and ears and lower temperatures on the back and tail.[47,48] For rats that were more active at night, when the light was turned off, the chance of the thermal lens detecting the areas around the eyes and ears may have increased, thereby increasing the average temperature. However, this finding was not confirmed in this study.

Table 2. Analysis of the F1 scores of the RD model based on the test set

	Test set 1		Test set 2		Test set 3		Test set 4		All test sets	
Class	Instances	F1 score	Instances	F1 score	Instances	F1 score	Instances	F1 score	Instances	F1 score
Drinking	117	0.950	124	0.912	58	0.941	52	0.953	351	0.936
Eating	223	0.854	103	0.981	98	0.953	109	0.954	533	0.919
Grooming	274	0.793	313	0.432	178	0.894	225	0.676	990	0.698
Rearing	572	0.939	215	0.986	263	0.998	309	0.990	1359	0.970
Gnawing	233	0.553	17	0.385	52	0.468	35	0.547	337	0.534
Searching	2125	0.903	1491	0.800	1737	0.941	1813	0.853	7166	0.879
Resting	28	0.409	1303	0.851	1170	0.960	1038	0.812	3539	0.863
Scratching	28	0.441	34	0.000	44	0.128	19	0.345	125	0.251
All class	3600	0.730	3600	0.668	3600	0.785	3600	0.766	14400	0.756

The recognition performance of the model was evaluated based on the four test subsets. The F1 scores of the grooming category ranged from 0.432 to 0.894 across test sets 1, 2, 3, and 4. The observed variation in values may be attributable to the diversity of grooming behaviors or individual differences among rats. The F1 scores of the resting category ranged from 0.409 to 0.960 across test sets 1, 2, 3, and 4. The variation in scores may be influenced by the difference number of images in test set 1.

In studies related to the SPT, Verharen et al.[40] studied the impact of sucrose preference on mouse licking behavior. The results revealed that the average degree of sucrose preference was 90.5%, and Liu et al.,[49] in work related to chronic mild stress (CMS) research, reported that sucrose preference ranged from 70 to 80%. According to our experimental results, rats spent significantly more time drinking water from bottles containing sucrose than from bottles without sucrose. Additionally, the sucrose preference was 94.3%, as shown in Figure 11C, which confirmed the rats' preference for sweet substances. This experiment demonstrated the application potential of the proposed system in sucrose preference testing. The time spent drinking from the two water bottles was determined by identifying the drinking behavior and location of the rats, and this innovative approach provides a fast, accurate, and objective analysis method for behavioral and neuroscience research. In conclusion, these two animal experiments verified the reliability and usability of the RW system for use in rat behavioral experiments.

Position tracking performance

The location tracking performance of a model is affected by various factors, and well-designed hardware settings are required to achieve optimal performance. Using a camera for position tracking is a simple and accurate method. However, when a camera is installed at a low position in the cage, the camera view could easily be covered by bedding or blocked by nest materials due to animal digging behavior. The multifunctional camera we designed was installed higher in the cage. The camera view was not obscured by environmental enrichment objects or nest materials in the cage, and a complete view of the cage could be continuously obtained. In addition, when animals hide in objects or nesting materials, part of the body view is blocked and cannot be identified. The thermal lens used in this work effectively detected the positions and temperatures of the animals through the gaps and heat conduction of the nest material. In the light cycle change experiment, a thermal lens was used to track the rat's position for 48 h, allowing a complete understanding of the rat's position within the cage, as shown in Figure 10G. The thermal lens data could still be used for position tracking without a complete view of the rat, thereby verifying the distribution of the nocturnal activities of the rats in the cage. We applied the thermal lens in a home cage with environmental enrichment objects to make the RW system robust for tracking rats.

Impact of dataset independence on performance

The training set is used to train the model in the model-building process, the validation set is used to fine-tune the parameters and prevent overfitting, and the test set is used to evaluate the model's performance with unseen data. Ensuring the independence and representativeness of these datasets is critical for the practical applicability of the model. Maleki et al.[50] verified that data samples from the same individual may be randomly assigned to the training and validation sets, resulting in data leakage. Data samples of the same individuals should not overlap in the training, validation, or test sets to prevent data leakage and maintain data independence.[48,50]

The training, validation, and test sets containing rat data used in this study were all extracted from videos. Behavior is typically characterized by a series of continuous movements. When manually selecting images to form a dataset, there was a tendency to densely oversample the same rat, leading to selected images that were too similar and correlated. Randomly dividing the dataset into training and validation sets for deep learning tasks is a suitable approach for large-scale datasets with sufficient diversity. When we randomly divided the dataset into training and validation sets at a ratio of 8:2, the mAPs of all the categories exceeded 0.9. The high accuracy was attributed to the correlation between the datasets. Therefore, ensuring that the same data do not appear in the training and validation sets helps to prevent overly high accuracy, which may impact the model's generalizability in practical scenarios, which may involve small datasets or datasets from a few

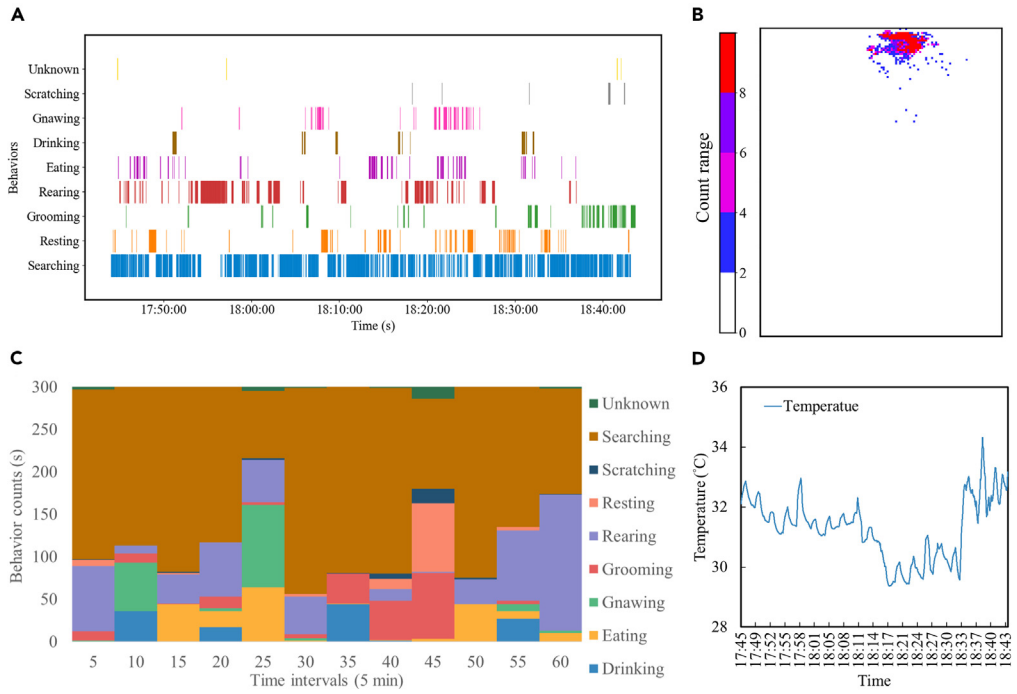


Figure 9. Test set 1 was analyzed to generate typical graphical representations

(A) Fence chart. Each category of rat behavior was plotted at 1-s intervals to present the occurrence time and duration of each behavior (See also Video S1).

(B) Heatmap. This heatmap can be used to determine a rat's preferred location or specific activity area.

(C) Stacked bar chart. For each behavioral category, the number of times each behavior occurred every 5 min was counted to evaluate the pattern and time distribution of the rat behavior.

(D) Body surface temperature change chart. The moving average curve of the body surface temperature data detected by the thermal lens was used to understand body temperature changes over time.

individuals. The validation sets used for deep learning tasks are often sampled randomly and widely. In this study, we partitioned the training and validation sets on the basis of samples collected during different periods. Since the model utilizes the modified NMS strategy, the resting behavior over continuous frames is considered in the recognition process. Therefore, the validation set consists of continuous images extracted from two 1-h videos. However, the RW system developed in this study, which was trained with a small-scale dataset, achieved great accuracy and was successfully applied in practical experiments.

Evaluation of the impact of imbalanced datasets

Imbalanced datasets are challenging for constructing deep learning models. Some classes have far more samples than other classes do, which is a common and expected problem in the real world. Similarly, this problem arose with the training and validation sets used in this study. The dataset included significantly more samples in the searching category than in the other categories; as a result, the model tended to predict searching behavior, as searching was the majority category. Undersampling is often used to balance datasets by removing samples from categories with the most data so that the number of samples is similar among all the categories, thus improving model performance. However, this method may lead to the loss of important feature information, ultimately affecting model accuracy. In our case, we attempted to use undersampling to process the searching category and rebuild the model based on the undersampled dataset; however, this led to a decrease in accuracy based on the validation set. Consequently, we decided to retain many samples in the searching category and not change the number of samples in the training set. Addressing data imbalance is a complex task. For categories that occur less frequently in the real world, more time and effort are required to collect sufficient representative samples. This is crucial for establishing a comprehensive and high-quality training set.

The advantages of the RW system

We comprehensively compared the advantages and characteristics of the RW system with those of other behavior recognition systems, as shown in Table 3. The first advantage is that the RW system uses only deep learning technology to identify rat behavior in real time without the use of any sensors. Moreover, real-time location tracking and body temperature detection can be achieved with the RW system. The second advantage is its suitability for installation in a standard conventional cage. The rats can be monitored in the home cage on the original

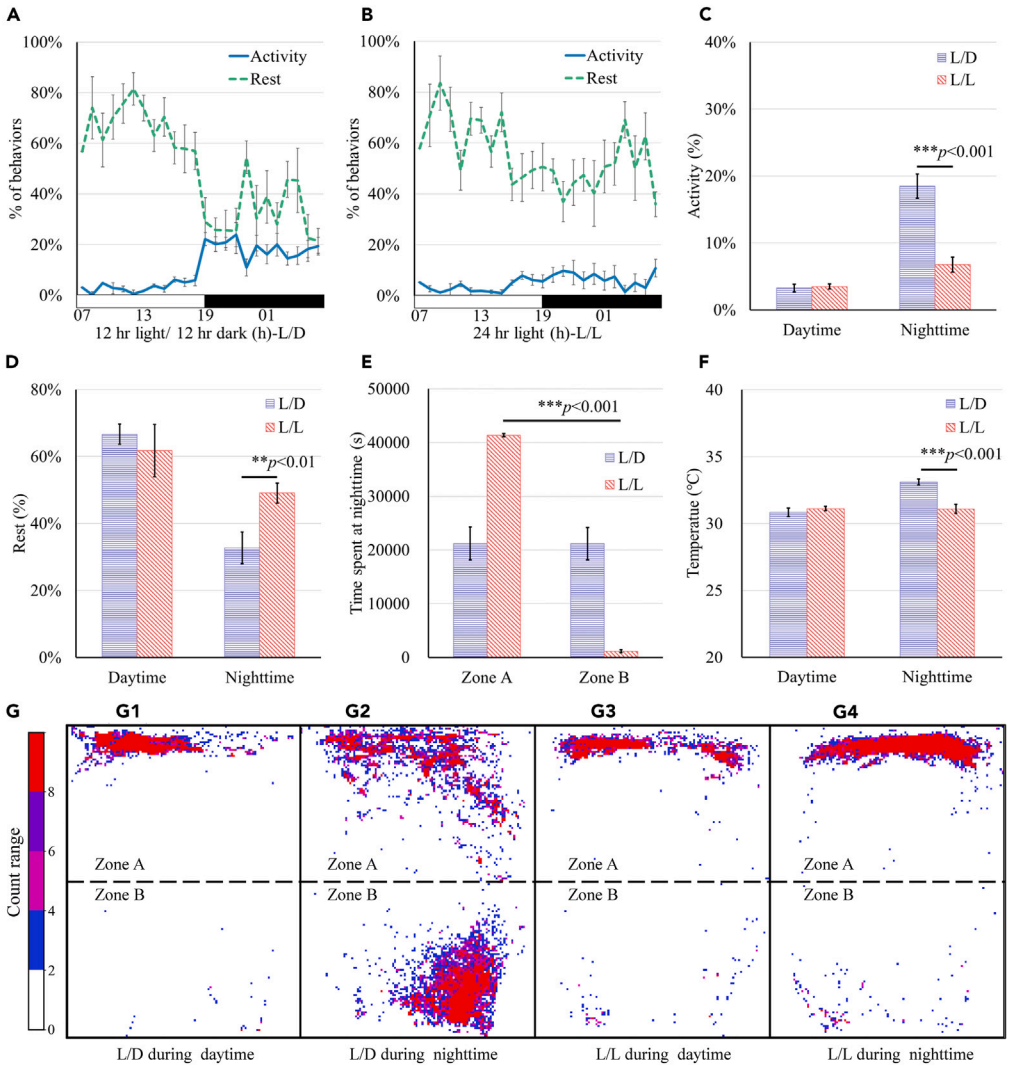


Figure 10. Experiments to determine the effects of light cycle changes on rats

(A) Activity and rest distribution of the L/D group under a normal light cycle. The main behavior exhibited by the rats during the day was resting, and their activity frequency was low, while the rats were more active at night. Data are represented as mean ± SEM.

(B) Activity and rest distribution of the L/L group under 24 h of light. Compared with that of the rats in the L/D group, the activity frequency of the rats in the L/L group decreased significantly due to nighttime light exposure. Data are represented as mean ± SEM.

(C) Comparison of the percentage of active time during the day and night. For the rats in the L/L and L/D groups, which had different light cycle conditions, the rats were active in the daytime less than 3.5 ± 0.4% of the time. At night, the rats in the L/D group were active 18.5 ± 1.8% of the time, which was significantly greater than the activity level of the rats in the L/L group (6.8 ± 1.1%)($p = 9.90 \times 10^{-5}$, *t* test, $n = 7$). Data are represented as mean ± SEM.

(D) Comparison of the percentage of rest time during the day and night. There was no significant difference in the percentage of rest time between the L/L and L/D groups during the 12 h of daytime lighting. However, the percentage of rest time in the L/L group when the lights were turned on at night was significantly greater than that in the L/D group. Data are represented as mean ± SEM.

(E) Comparison of the location distribution within the cages at night with the lights turned off. The space in the housing cage was divided into two zones. Zone A was close to the wall side, and the brightness was lower because the area was shielded. Zone B represents the area close to the outside of the rack, which is brighter. There was no significant difference in the time that the rats in the L/D group spent in zone A or zone B ($p = 0.99$, *t* test, $n = 7$). However, the rats in the L/L group, which were exposed to light at night, spent significantly more time in zone A to reduce their exposure to light ($p = 4.65 \times 10^{-10}$, *t* test, $n = 7$). Data are represented as mean ± SEM.

(F) Comparison of body temperature changes during the day and night. There was no significant difference in the daytime average body surface temperature between the L/D and L/L groups ($p = 0.19$, *t* test, $n = 7$). The average body surface temperature of the rats in the L/D group was significantly greater than that of the rats in the L/L group at night ($p = 6.31 \times 10^{-5}$, *t* test, $n = 7$), indicating that the body temperature of the rats was greater at night than during the day under a normal light cycle. Data are represented as mean ± SEM.

(G) Heatmap of the position distribution of the rats affected by brightness in the housing cage.

Figure 10. *Continued*

(G1 and G3) When the light was turned on during the day, the rats in the L/D group and L/L group preferred zone A of the housing cage and stayed as close to the wall side as possible to reduce exposure to light.

(G2) When the light was turned off at night, the rats in the L/D group had footprints distributed in both zone A and zone B.

(G4) When the nighttime light was turned on, the rats in the L/L group preferred zone A to reduce their exposure to light.

cage rack without the need to move the home cage. This reduces cage costs and time needed for animal adaptation. This design is more convenient for researchers, whereas conventional systems have complex multisensor designs and space-consuming arrangements. The third advantage is the physiological monitoring of body surface temperature with the RW system. Real-time monitoring of rat body temperature changes and behavioral responses helps researchers gain a more comprehensive understanding of the physiological states of rats. The fourth advantage is that the RW system uses only a single multifunction camera that is easy to install. The camera is designed to be lightweight, portable, space-saving, and highly scalable. This approach can easily be scaled up and has better space utilization than conventional systems, which is advantageous for conducting high-throughput experiments.

Additionally, we compared the accuracy of the RW system with that of seven behavior recognition systems, as presented in Table 4. Each system can identify various behavior categories, utilizing a variety of evaluation metrics such as the F1 score and confusion matrix. While direct numerical comparisons may be biased due to these discrepancies, the results still offer valuable insights. The F1 score is a metric that balances precision and recall. The value of each cell in the confusion matrix represents the proportion of instances in each true class that are classified into a given predicted class. The F1 scores and confusion matrix values of the RW system for drinking, eating, and rearing behaviors ranged from 0.92 to 0.97 and 0.89 to 0.96, respectively. In contrast, the F1 scores and confusion matrix values of the other four systems were between 0.76 and 0.95 and between 0.42 and 0.79, respectively. This demonstrates the good performance of the RW system and its distinct advantages in recognizing these behaviors. In terms of grooming behavior, the F1 score of the RW system was 0.70, which is comparable to the F1 score of 0.70 reported in a 2020 study. However, the confusion matrix value of 0.56 was lower than the range of 0.70–0.74 obtained with the other systems. This suggests that our system may require further optimization under specific conditions. Resting or sleep behavior accounts for a substantial portion of daily activities of rodents and is generally well detected by most recognition systems. The RW system demonstrated an F1 score of 0.86 for resting behavior, with a confusion matrix value of 0.94. The other systems presented F1 scores ranging from 0.85 to 0.87 and confusion matrix values between 0.85 and 0.94, indicating that their performance in this category was comparable to that of the RW system. Searching and sniffing behaviors were combined into a single category for cross-system comparison. The RW system achieved F1 scores and confusion matrix values of 0.88 and 0.87, respectively, for this category. These results exceed the accuracy range of 0.66–0.70 observed with the other two systems. The RW system does not include a distinct category for walking behavior. Instead, walking behavior is included in the default searching category. The four recognition systems in the comparison table all include walking categories, with accuracies ranging from 0.55 to 0.86. This suggests that walking behavior is a category that could be included in the RW system in the future. The accuracy of the RW system in detecting gnawing and scratching behaviors was suboptimal. Since these behaviors were not included in the recognition tasks of other systems, a direct comparison was not feasible. In conclusion, despite the use of different evaluation metrics across systems, our system demonstrated outstanding performance in identifying behaviors, particularly drinking and eating. This comparison highlights the strengths of the RW system and identifies potential areas for improvement, providing valuable insights for future research.

In summary, the RW system has several notable advantages, demonstrating its value in rat behavior research. Its use of a single multifunctional camera, coupled with integrated physiological monitoring systems, offers a streamlined and space-efficient solution that is readily deployable in standard conventional cages, without the need for additional sensors or complex installations. Furthermore, the RW system demonstrated superior accuracy in the recognition of key behaviors such as drinking, eating, and rearing. These strengths demonstrate that the RW system is a highly promising research tool with significant potential to advance rodent behavioral studies.

Future work

The F1 scores of some categories based on the validation and test sets were low. The reason may be that the number of data samples is too small, the near-infrared image quality is poor, or the image blur limits the judgment ability of the human annotator. In future work, the use of high-performance edge devices to increase data sample sizes and acquire high-resolution images could support the development of large-scale models, potentially resulting in a significant improvement in accuracy. Additionally, transitioning from wired to wireless data transmission could increase the convenience of installing the system, representing a potential research direction. Photoacoustic technology has been widely used in preclinical research, including detecting metastatic tumors,[58] evaluating the correlation between neural activity and hemodynamics,[59,60] and performing functional assessments of arthritis.[61] However, studies have shown differences in hemodynamic and neurological responses between anesthetized and freely moving rats.[62–64] Thus, wearable photoacoustic devices could be integrated into the RW system to reveal the correlation between neural activity and behavioral changes in freely moving rats; we are currently developing such a system.

Conclusion

In this study, an RW system suitable for application in a familiar low-pressure home cage environment was developed. The RW system is a multifunctional design created by using 3D printing and an edge device that enables real-time rat behavior recognition, location tracking, and body temperature change measurements via a GUI tool. The improved recognition model was constructed on the basis of the YOLOv5s neural network architecture. In this process, we used contextual object labeling and a modified NMS scheme to increase the

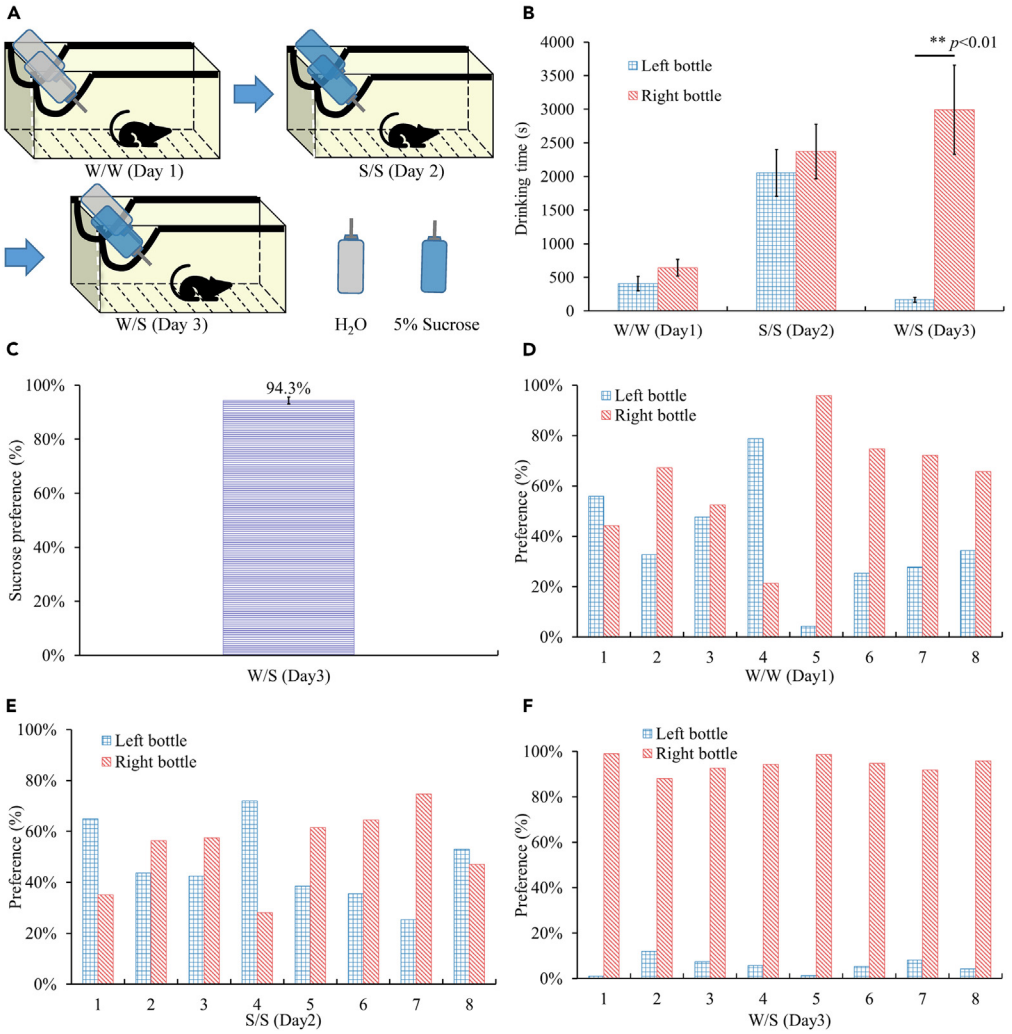


Figure 11. Analysis of the sucrose preference test results

(A) Schematic diagram of the experimental procedure. The W/W group was given two bottles of sucrose-free water on the first day, and the S/S group was given two bottles of 5% sucrose-containing water on the second day. The W/S group was given one sucrose-free water bottle and one sucrose-containing water bottle on the third day to test the preferences of the rats.

(B) Comparison of the time spent drinking water from the left and right water bottles by the rats in each group. The drinking category was identified by the RW system every second according to the position of the rat relative to the positions of the two water bottles, and the time spent drinking water from the water bottle was obtained. There was no significant difference in drinking time between the two sucrose-free water bottles in the W/W group ($p = 0.48$, t test, $n = 8$). The time spent drinking water containing sucrose from the two bottles in the S/S group was greater than that in the W/W group. The time spent drinking water from the right sucrose-containing water bottle in the W/S group was significantly greater than that spent drinking from the left sucrose-free water bottle ($p = 3.56 \times 10^{-3}$, t test, $n = 8$). Data are represented as mean ± SEM.

(C) The sucrose preference of the rats in the W/S group was 94.3 ± 1.3%. The results showed that the rats clearly preferred sucrose. Data are represented as mean ± SEM.

(D and E) Analysis of water bottle preference of individual rats in the W/W group and S/S group. Individual rats may have different preferences for the left and right water bottles, but no consistent preferences among individual rats were observed.

(F) Analysis of sucrose preference of individual rats in the W/S group. Individual rats showed a preference for sucrose by consistently choosing the right water bottle containing sucrose.

accuracy of detecting the drinking and resting categories. The accuracy of the model based on the test set and the F1 scores of the drinking, eating, rearing, searching, and resting categories were greater than 0.8, with the F1 scores for the drinking, eating, and rearing categories exceeding 0.9. The average F1 score across all categories was 0.756. We conducted two experiments to verify the practicality of the RW system in behavioral research. In the light cycle change experiment, the rats in the L/D group, with the lights turned off at night, were active 18.5 ± 1.8% of the time, which was significantly greater than the 6.8 ± 1.1% value in the L/L group ($p = 9.90 \times 10^{-5}$, t test, $n = 7$). These results

Table 3. Comparison of characteristics between the RW system and other behavior recognition systems

Reference	This work	2018[51]	2020[52]	2013[53]	2010[54]	2019[55]	2012[56]	2016[57]
Classes of behaviors	8	9	4 or 7	10	8	2[a]	2[a]	1[a]
Classification technology	YOLOv5	SVM with RBF kernel	SVM	Image processing, ML	SVM-HMM	Image processing	Sensors, infrared	RFID, infrared
Real-time behavior recognition	Real-time	Offline	Offline	Near real-time	Offline	Offline	Real-time	Offline
Operational Space	Small footprint	Large footprint	Large footprint	Large footprint	Small footprint	Small footprint	Small footprint	Small footprint
Home-cage Design	Conventional cage	Conventional cage	Open-field	Proprietary cage	Conventional cage	Proprietary cage	Conventional cage	Conventional cage
Detection quantity of animals	Single	Single	Single	Single	Single	Single	Single	Multiple
Camera perspective	Front view	Top view	Top view	Top view	Side view	Top view	–	Top view and side view
Tracking detection	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Body temperature detection	Yes	No	No	No	No	No	No	No

ML: Machine learning.
The main differences between the RW system and other behavioral recognition systems included the recognition technology, system functions, and operating space.[51–57] The RW system uses only deep learning technology to identify rat behavior in real time without the use of any sensors; in addition, the RW system can track location and body temperature simultaneously. Moreover, the RW system can easily be installed in a standard conventional cage, thus reducing costs and adaptation time. Additionally, body surface temperature and behavioral responses can be monitored with the RW system, providing comprehensive physiological insights. The single multifunction camera is easy to install, portable, and scalable, making it ideal for high-throughput experiments.
[a]Application scenarios can broaden behavioral categories.

are consistent with the nocturnal habits of rodents. In the SPT, the times spent drinking from the two water bottles were obtained separately by analyzing drinking behavior and location data. The results revealed that the sucrose preference was 94.3 $\pm$ 1.3%, indicating that the rats clearly preferred sucrose, which is consistent with the known preference of rats for sweet substances. Overall, this research represents an important breakthrough in rat behavior recognition and provides approaches for similar behavior recognition tasks. The RW system is expected to become a research tool with great application potential.

Limitations of the study

A multifunctional camera is installed in the cage to obtain a complete image of the cage at a wide angle of approximately 45°. However, it is difficult for humans to determine a rat's behavior category when the rat is positioned below or facing away from the camera. During the experiment, the use of a single multifunctional camera was sufficient to determine the rats' behavioral categories and body temperatures. However, some behavioral experiments require identifying rat behavior in all directions to obtain comprehensive behavioral data without omissions. Multiple cameras must be deployed to address this issue. Owing to the social nature of rats, housing individual rats in single cages should be avoided unless required by the experimental design. Currently, the RW system is applicable only in single-housing experiments. However, recognizing the behaviors of multiple rats is challenging because of the complex social behaviors exhibited by rats. The increased activity of multiple rats in the same enclosure results in greater occlusion within the FOV, complicating model identification and tracking. Incorporating multiple cameras from different angles could mitigate these issues.

The RW system is currently suitable for conventional rat housing cages of standard sizes. Adapting the system to different housing configurations would require adjustments to the camera's FOV and positioning to maintain similar heights and angles. Additionally, if other housing systems use feeders or water bottles with different appearances, the behavior identification accuracy will be affected. In addition, to apply the RW system to rats with different fur colors, the amount of training set data should be increased to improve the model's generalizability. However, for mice with large differences in appearance and size, a new dataset specifically for the relevant mice would be needed to develop a suitable model. Therefore, future work should focus on developing an extended version of the system that incorporates multiple cameras. Additional research objectives include creating models capable of recognizing multiple individual rats and expanding the system to other species. Moreover, a meticulous labeling process is crucial, as errors due to differences between observers or laboratories can occur.[9–11] In this study, the training, validation, and test sets were all annotated by the same researcher. While this method guarantees consistency in labeling judgments, it also introduces a potential systemic bias in model predictions. To address this issue, multiple researchers could annotate the dataset according to a defined ethogram in future research. Statistical analysis of interobserver variability could be performed

Table 4. Comparison of the accuracy of the RW system and other behavioral recognition systems

Reference	This work		2018[51]	2020[52]	2013[53]	2010[54]	2019[55]	2012[56]	2016[57]
Evaluation metric	F1 score	Confusion matrix	F1 score[a]	F1 score (4 classes)	Confusion matrix	Confusion matrix	Custom metric	CageScore™	Spearman's rank coefficient
Drinking	0.94	0.94	–	–	0.42	0.72	–	–	–
Eating	0.92	0.89	–	–	0.79	0.75	–	–	–
Grooming	0.70	0.56	0.60	70.3%	0.74	0.70	–	–	–
Sup. Rearing/Uns. Rearing	0.97	0.96	0.85/0.76	95.1%	0.70	0.70	–	–	–
Gnawing	0.53	0.49	–	–	–	–	–	–	–
Searching/ Sniffing	0.88	0.87	0.66	–	0.70	–	–	–	–
Resting/Sleep	0.86	0.94	0.85	87.3%	0.00	0.94	90.5%	85-90%	–
Scratching	0.25	0.17	–	–	–	–	–	–	–
Walking	–	–	0.68	70.9%	0.86	0.55	–	–	–
Hang	–	–	–	–	–	0.92	–	–	–
Bending	–	–	0.81	–	–	–	–	–	–
Micro-motion	–	–	0.70	–	–	0.83	–	–	–
Twitch	–	–	–	–	0.81	–	–	–	–
Activity	–	–	–	–	–	–	–	–	0.95

The RW system demonstrated excellent performance in detecting drinking, eating, and rearing behaviors, achieving F1 scores between 0.92 and 0.97 and confusion matrix values ranging from 0.89 to 0.96. In comparison, for other systems, the F1 scores ranged from 0.76 to 0.95, and the confusion matrix values ranged from 0.42 to 0.79. For grooming behavior, the F1 score of the RW system was 0.70, which is consistent with values reported for other systems. However, its confusion matrix value of 0.56 is lower than the 0.70 to 0.74 range obtained with other systems. With respect to resting or sleeping behavior, the performance of the RW system was comparable to that of the other systems, with an F1 score of 0.86 and a confusion matrix value of 0.94. The F1 scores of the other systems ranged from 0.85 to 0.87, and the confusion matrix values were between 0.85 and 0.94. For detecting searching and sniffing behaviors, the RW system achieved an F1 score of 0.88 and a confusion matrix value of 0.87, surpassing the 0.66 to 0.70 accuracy range reported by other systems. The detection of gnawing and scratching behaviors by the RW system was suboptimal, and direct comparisons with other systems were not possible because these behaviors were excluded from their recognition tasks.[51–57]

[a]F1 score is the average of two test sets.

to select appropriate annotators for establishing the ground truth labels.[18] Alternatively, in cases of discrepancies among annotators, a consensus model can be created by combining the labels from the best annotators on the basis of the majority vote.[65] The incorporation of these approaches will enhance the objectivity and reliability of annotations in future studies.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, L.-D.L. (ldliao@nhri.edu.tw).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data reported in this article will be shared by the lead contact upon request.
- All original code has been deposited at GitHub and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any addition information required to reanalyze the data reported in this paper is available from the lead contact L.-D.L. (ldliao@nhri.edu.tw) upon request.

ACKNOWLEDGMENTS

This research was supported in part by the National Science and Technology Council of Taiwan under grant numbers 110-2221-E-400-003-MY3, 111-3114-8-400-001, 111-2314-B-075-006, 111-2221-E-035-015, 111-2218-E-007-019, and 112-2629-E-400-001; the National Health Research Institutes of Taiwan under grant numbers NHRI-EX111-11111EI and NHRI-EX111-11129EI; the Ministry of Health and Welfare, Taiwan, under grant numbers MOHW 112-0324-01-30-06 and MOHW 113-0324-01-30-11; and the Metal Industries Research & Development Centre under grant number 112-EC-17-A-22-1851.

AUTHOR CONTRIBUTIONS

Conceptualization, L.-D.L. and C.-M.H.; Methodology, L.-D.L. and C.-M.H.; Validation, L.-D.L. and C.-M.H.; Investigation, C.-M.H.; Resources, L.-D.L.; Data Curation, L.-D.L. and C.-M.H.; Writing – Original Draft, C.-M.H.; Writing – Review and Editing, C.-M.H., C.-H.H., J.-K.C., and L.-D.L.; Funding Acquisition, C.-H.H., J.-K.C., and L.-D.L.

DECLARATION OF INTERESTS

The authors declare no conflicts of interest.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Hardware of the multifunctional camera
  - Program for determining location and body surface temperature
  - Dataset structure and sources
  - Model training
  - Contextual object labeling
  - Modification of nonmaximum suppression
  - Evaluation metrics
  - Animal experimental research design
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111223.

Received: April 15, 2024
Revised: June 13, 2024
Accepted: October 18, 2024
Published: October 21, 2024

REFERENCES

1. Shi, X., Bai, H., Wang, J., Wang, J., Huang, L., He, M., Zheng, X., Duan, Z., Chen, D., Zhang, J., et al. (2021). Behavioral Assessment of Sensory, Motor, Emotion, and Cognition in Rodent Models of Intracerebral Hemorrhage. Front. Neurol. *12*, 667511. https://doi.org/10.3389/fneur.2021.667511.
2. Himanshu; Dharmila, and Sarkar, D.; Nutan (2020). A Review of Behavioral Tests to Evaluate Different Types of Anxiety and Anti-anxiety Effects. Clin. Psychopharmacol. Neurosci. *18*, 341–351. https://doi.org/10.9758/cpn.2020.18.3.341.
3. Balkaya, M., Kröber, J.M., Rex, A., and Endres, M. (2013). Assessing post-stroke behavior in mouse models of focal ischemia. J. Cereb. Blood Flow Metab. *33*, 330–338. https://doi.org/10.1038/jcbfm.2012.185.
4. Bruchas, M.R., Land, B.B., and Chavkin, C. (2010). The dynorphin/kappa opioid system as a modulator of stress-induced and pro-addictive behaviors. Brain Res. *1314*, 44–55. https://doi.org/10.1016/j.brainres.2009.08.062.
5. Gobira, P.H., Ropke, J., Aguiar, D.C., Crippa, J.A.S., and Moreira, F.A. (2013). Animal models for predicting the efficacy and side effects of antipsychotic drugs. Braz. J. Psychiatry. *35*, S132–S139. https://doi.org/10.1590/1516-4446-2013-1164.
6. Drai, D., Benjamini, Y., and Golani, I. (2000). Statistical discrimination of natural modes of motion in rat exploratory behavior. J. Neurosci. Methods *96*, 119–131. https://doi.org/10.1016/s0165-0270(99)00194-6.
7. Hammond, R.S., Tull, L.E., and Stackman, R.W. (2004). On the delay-dependent involvement of the hippocampus in object recognition memory. Neurobiol. Learn. Mem. *82*, 26–34. https://doi.org/10.1016/j.nlm.2004.03.005.
8. Murai, T., Okuda, S., Tanaka, T., and Ohta, H. (2007). Characteristics of object location memory in mice: Behavioral and pharmacological studies. Physiol. Behav. *90*, 116–124. https://doi.org/10.1016/j.physbeh.2006.09.013.
9. Crabbe, J.C., Wahlsten, D., and Dudek, B.C. (1999). Genetics of mouse behavior: interactions with laboratory environment. Science *284*, 1670–1672. https://doi.org/10.1126/science.284.5420.1670.
10. Wahlsten, D., Metten, P., Phillips, T.J., Boehm, S.L., 2nd, Burkhart-Kasch, S., Dorow, J., Doerksen, S., Downing, C., Fogarty, J., Rodd-Henricks, K., et al. (2003). Different data from different labs: lessons from studies of gene-environment interaction. J. Neurobiol. *54*, 283–311. https://doi.org/10.1002/neu.10173.
11. Sare, R.M., Lemons, A., and Smith, C.B. (2021). Behavior Testing in Rodents: Highlighting Potential Confounds Affecting Variability and Reproducibility. Brain Sci. *11*, 522. https://doi.org/10.3390/brainsci11040522.
12. Klein, C.J.M.I., Budiman, T., Homberg, J.R., Verma, D., Keijer, J., and van Schothorst, E.M. (2022). Corrigendum: Measuring Locomotor Activity and Behavioral Aspects of Rodents Living in the Home-Cage. Front. Behav. Neurosci. *16*, 943307. https://doi.org/10.3389/fnbeh.2022.943307.
13. Spruijt, B.M., and DeVisser, L. (2006). Advanced behavioural screening: automated home cage ethology. Drug Discov. Today Technol. *3*, 231–237. https://doi.org/10.1016/j.ddtec.2006.06.010.
14. Kahnau, P., Mieske, P., Wilzopolski, J., Kalliokoski, O., Mandillo, S., Hölter, S.M., Voikar, V., Amfim, A., Badurek, S., Bartelik, A., et al. (2023). A systematic review of the development and application of home cage monitoring in laboratory mice and rats. BMC Biol. *21*, 256. https://doi.org/10.1186/s12915-023-01751-7.
15. Kafkafi, N., Agassi, J., Chesler, E.J., Crabbe, J.C., Crusio, W.E., Eilam, D., Gerlai, R., Golani, I., Gomez-Marin, A., Heller, R., et al. (2018). Reproducibility and replicability of rodent phenotyping in preclinical studies. Neurosci. Biobehav. Rev. *87*, 218–232. https://doi.org/10.1016/j.neubiorev.2018.01.003.
16. Krynitsky, J., Legaria, A.A., Pai, J.J., Garmendia-Cedillos, M., Salem, G., Pohida, T., and Kravitz, A.V. (2020). Rodent Arena Tracker (RAT): A Machine Vision Rodent Tracking Camera and Closed Loop Control System. eNeuro *7*, ENEURO.0485-19.2020. https://doi.org/10.1523/ENEURO.0485-19.2020.
17. Grieco, F., Bernstein, B.J., Biemans, B., Bikovski, L., Burnett, C.J., Cushman, J.D., van Dam, E.A., Fry, S.A., Richmond-Hacham, B., Homberg, J.R., et al. (2021). Measuring Behavior in the Home Cage: Study Design,

Applications, Challenges, and Perspectives. Front. Behav. Neurosci. *15*, 735387. https://doi.org/10.3389/fnbeh.2021.735387.

18. Segalin, C., Williams, J., Karigo, T., Hui, M., Zelikowsky, M., Sun, J.J., Perona, P., Anderson, D.J., and Kennedy, A. (2021). The Mouse Action Recognition System (MARS) software pipeline for automated analysis of social behaviors in mice. Elife *10*, e63720. https://doi.org/10.7554/eLife.63720.
19. Redfern, W.S., Tse, K., Grant, C., Keerie, A., Simpson, D.J., Pedersen, J.C., Rimmer, V., Leslie, L., Klein, S.K., Karp, N.A., et al. (2017). Automated recording of home cage activity and temperature of individual rats housed in social groups: The Rodent Big Brother project. PLoS One *12*, e0181068. https://doi.org/10.1371/journal.pone.0181068.
20. Bains, R.S., Wells, S., Sillito, R.R., Armstrong, J.D., Cater, H.L., Banks, G., and Nolan, P.M. (2018). Assessing mouse behaviour throughout the light/dark cycle using automated in-cage analysis tools. J. Neurosci. Methods *300*, 37–47. https://doi.org/10.1016/j.jneumeth.2017.04.014.
21. Datta, S.R., Anderson, D.J., Branson, K., Perona, P., and Leifer, A. (2019). Computational Neuroethology: A Call to Action. Neuron *104*, 11–24. https://doi.org/10.1016/j.neuron.2019.09.038.
22. 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., and Harvey, C.D. (2021). DeepEthogram, a machine learning pipeline for supervised behavior classification from raw pixels. Elife *10*, e63377. https://doi.org/10.7554/eLife.63377.
23. Pereira, T.D., Aldarondo, D.E., Willmore, L., Kislin, M., Wang, S.S.H., Murthy, M., and Shaevitz, J.W. (2019). Fast animal pose estimation using deep neural networks. Nat. Methods *16*, 117–125. https://doi.org/10.1038/s41592-018-0234-5.
24. de Chaumont, F., Lemière, N., Coqueran, S., Bourgeron, T., and Ey, E. (2021). LMT USV Toolbox, a Novel Methodological Approach to Place Mouse Ultrasonic Vocalizations in Their Behavioral Contexts-A Study in Female and Male C57BL/6J Mice and in Shank3 Mutant Females. Front. Behav. Neurosci. *15*, 735920. https://doi.org/10.3389/fnbeh.2021.735920.
25. Liu, W.L., Wang, Y., Chen, Y.X., Chen, B.Y., Lin, A.Y.C., Dai, S.T., Chen, C.H., and Liao, L.D. (2023). An IoT-based smart mosquito trap system embedded with real-time mosquito image processing by neural networks for mosquito surveillance. Front. Bioeng. Biotechnol. *11*, 1100968. https://doi.org/10.3389/fbioe.2023.1100968.
26. Yu, L., Guo, J., Pu, Y., Cen, H., Li, J., Liu, S., Nie, J., Ge, J., Yang, S., Zhao, H., et al. (2023). A Recognition Method of Ewe Estrus Crawling Behavior Based on Multi-Target Detection Layer Neural Network. Animals. *13*, 413. https://doi.org/10.3390/ani13030413.
27. Girardie, O., Bonneau, M., Billon, Y., Bailly, J., David, I., and Canario, L. (2022). Analysis of image-based sow activity patterns reveals several associations with piglet survival and early growth. Front. Vet. Sci. *9*, 1051284. https://doi.org/10.3389/fvets.2022.1051284.
28. Jocher, G. (2020). YOLOv5 by Ultralytics. https://github.com/ultralytics/yolov5.
29. Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Michael, K., Fang, J., Yifu, Z., Wong, C., and Montes, D. (2022). ultralytics/yolov5: V7. 0-yolov5 Sota Realtime Instance Segmentation. Zenodo.
30. Taylor, K., and Alvarez, L.R. (2019). An Estimate of the Number of Animals Used for Scientific Purposes Worldwide in 2015. Altern. Lab. Anim. *47*, 196–213. https://doi.org/10.1177/0261192919899853.
31. Carbone, L. (2021). Estimating mouse and rat use in American laboratories by extrapolation from Animal Welfare Act-regulated species. Sci. Rep. *11*, 493. https://doi.org/10.1038/s41598-020-79961-0.
32. European Commission (2010). Directive 2010/63/EU on the Protection of Animals Used for Scientific Purposes (Official Journal of the European Union).
33. Neuhaus, W., Reininger-Gutmann, B., Rinner, B., Plasenzotti, R., Wilflingseder, D., De Kock, J., Vanhaecke, T., Rogiers, V., Jírová, D., Kejlová, K., et al. (2022). The Rise of Three Rs Centres and Platforms in Europe. Altern. Lab. Anim. *50*, 90–120. https://doi.org/10.1177/02611929221099165.
34. Wu, J.Y., Wang, Y., Ching, C.T.S., Wang, H.M.D., and Liao, L.D. (2023). IoT-based wearable health monitoring device and its validation for potential critical and emergency applications. Front. Public Health *11*, 1188304. https://doi.org/10.3389/fpubh.2023.1188304.
35. Al Bassam, N., Hussain, S.A., Al Qaraghuli, A., Khan, J., Sumesh, E.P., and Lavanya, V. (2021). IoT based wearable device to monitor the signs of quarantined remote patients of COVID-19. Inform. Med. Unlocked *24*, 100588. https://doi.org/10.1016/j.imu.2021.100588.
36. Channa, A., Popescu, N., Skibinska, J., and Burget, R. (2021). The Rise of Wearable Devices during the COVID-19 Pandemic: A Systematic Review. Sensors *21*, 5787. https://doi.org/10.3390/s21175787.
37. Wu, J.Y., Ching, C.T.S., Wang, H.M.D., and Liao, L.D. (2022). Emerging Wearable Biosensor Technologies for Stress Monitoring and Their Real-World Applications. Biosensors *12*, 1097. https://doi.org/10.3390/bios12121097.
38. Talaat, F.M., and El-Balka, R.M. (2023). Stress monitoring using wearable sensors: IoT techniques in medical field. Neural Comput. Appl. *2*, 1–14. https://doi.org/10.1007/s00521-023-08681-z.
39. Gomes, N., Pato, M., Lourenco, A.R., and Datia, N. (2023). A Survey on Wearable Sensors for Mental Health Monitoring. Sensors *23*, 1330. https://doi.org/10.3390/s23031330.
40. Verharen, J.P.H., de Jong, J.W., Zhu, Y., and Lammel, S. (2023). A computational analysis of mouse behavior in the sucrose preference test. Nat. Commun. *14*, 2419. https://doi.org/10.1038/s41467-023-38028-0.
41. Depres-Brummer, P., Levi, F., Metzger, G., and Touitou, Y. (1995). Light-induced suppression of the rat circadian system. Am. J. Physiol. *268*, R1111–R1116. https://doi.org/10.1152/ajpregu.1995.268.5.R1111.
42. Ikeda, M., Sagara, M., and Inoué, S. (2000). Continuous exposure to dim illumination uncouples temporal patterns of sleep, body temperature, locomotion and drinking behavior in the rat. Neurosci. Lett. *279*, 185–189. https://doi.org/10.1016/s0304-3940(99)00943-x.
43. Gutiérrez-Pérez, M., González-González, S., Estrada-Rodriguez, K.P., Espítia-Bautista, E., Guzmán-Ruiz, M.A., Escalona, R., Escobar, C., and Guerrero-Vargas, N.N. (2023). Dim Light at Night Promotes Circadian Disruption in Female Rats, at the Metabolic, Reproductive, and Behavioral Level. Adv Biol-Ger *7*, e2200289. https://doi.org/10.1002/adbi.202200289.
44. Kim, D., Hanzawa, F., Shimizu, H., Sun, S., Umeki, M., Ikeda, S., Mochizuki, S., and Oda, H. (2023). Delayed feeding of a high-sucrose diet led to increased body weight by affecting the circadian rhythm of body temperature and hepatic lipid-metabolism genes in rats. J. Nutr. Biochem. *111*, 109185. https://doi.org/10.1016/j.jnutbio.2022.109185.
45. Goh, G.H., Mark, P.J., and Maloney, S.K. (2016). Altered energy intake and the amplitude of the body temperature rhythm are associated with changes in phase, but not amplitude, of clock gene expression in the rat suprachiasmatic nucleus in vivo. Chronobiol. Int. *33*, 85–97. https://doi.org/10.3109/07420528.2015.1112395.
46. Studholme, K.M., Gompf, H.S., and Morin, L.P. (2013). Brief light stimulation during the mouse nocturnal activity phase simultaneously induces a decline in core temperature and locomotor activity followed by EEG-determined sleep. Am. J. Physiol. Regul. Integr. Comp. Physiol. *304*, R459–R471. https://doi.org/10.1152/ajpregu.00460.2012.
47. Vianna, D.M.L., and Carrive, P. (2005). Changes in cutaneous and body temperature during and after conditioned fear to context in the rat. Eur. J. Neurosci. *21*, 2505–2512. https://doi.org/10.1111/j.1460-9568.2005.04073.x.
48. Dominguez-Oliva, A., Hernandez-Avalos, I., Olmos-Hernandez, A., Villegas-Juache, J., Verduzco-Mendoza, A., and Mota-Rojas, D. (2023). Thermal Response of Laboratory Rats (Rattus norvegicus) during the Application of Six Methods of Euthanasia Assessed by Infrared Thermography. Animals (Basel) *13*, 2820. https://doi.org/10.3390/ani13182820.
49. Liu, M.Y., Yin, C.Y., Zhu, L.J., Zhu, X.H., Xu, C., Luo, C.X., Chen, H., Zhu, D.Y., and Zhou, Q.G. (2018). Sucrose preference test for measurement of stress-induced anhedonia in mice. Nat. Protoc. *13*, 1686–1698. https://doi.org/10.1038/s41596-018-0011-z.
50. Maleki, F., Ovens, K., Gupta, R., Reinhold, C., Spatz, A., and Forghani, R. (2023). Generalizability of Machine Learning Models: Quantitative Evaluation of Three Methodological Pitfalls. Radiol. Artif. Intell. *5*, e220028. https://doi.org/10.1148/ryai.220028.
51. Wang, Z., Mirbozorgi, S.A., and Ghovanloo, M. (2018). An automated behavior analysis system for freely moving rodents using depth image. Med. Biol. Eng. Comput. *56*, 1807–1821. https://doi.org/10.1007/s11517-018-1816-1.
52. Geros, A., Magalhaes, A., and Aguiar, P. (2020). Improved 3D tracking and automated classification of rodents' behavioral activity using depth-sensing cameras. Behav. Res. Methods *52*, 2156–2167. https://doi.org/10.3758/s13428-020-01381-9.
53. van Dam, E.A., van der Harst, J.E., ter Braak, C.J.F., Tegelenbosch, R.A.J., Spruijt, B.M., and Noldus, L.P.J.J. (2013). An automated system for the recognition of various specific rat behaviours. J. Neurosci. Methods *218*, 214–224. https://doi.org/10.1016/j.jneumeth.2013.05.012.
54. Jhuang, H., Garrote, E., Mutch, J., Yu, X., Khilnani, V., Poggio, T., Steele, A.D., and

Serre, T. (2010). Automated home-cage behavioural phenotyping of mice. Nat. Commun. *1*, 68. https://doi.org/10.1038/ncomms1064.

55. Singh, S., Bermudez-Contreras, E., Nazari, M., Sutherland, R.J., and Mohajerani, M.H. (2019). Low-cost solution for rodent home-cage behaviour monitoring. PLoS One *14*, e0220751. https://doi.org/10.1371/journal.pone.0220751.
56. Xie, X.S., Zhang, J., Zou, B., Xie, J., Fang, J., Zaveri, N.T., and Khroyan, T.V. (2012). Rodent Behavioral Assessment in the Home Cage Using the SmartCage™ System. Animal Models of Acute Neurological Injuries II. Injury Mechan. Assess. *1*, 205–222.
57. Bains, R.S., Cater, H.L., Sillito, R.R., Chartsias, A., Sneddon, D., Concas, D., Keskivali-Bond, P., Lukins, T.C., Wells, S., Acevedo Arozena, A., et al. (2016). Analysis of Individual Mouse Activity in Group Housed Animals of Different Inbred Strains using a Novel Automated Home Cage Analysis System. Front. Behav. Neurosci. *10*, 106. https://doi.org/10.3389/fnbeh.2016.00106.
58. Luke, G.P., Myers, J.N., Emelianov, S.Y., and Sokolov, K.V. (2014). Sentinel lymph node biopsy revisited: ultrasound-guided photoacoustic detection of micrometastases using molecularly targeted plasmonic nanosensors. Cancer Res. *74*, 5397–5408. https://doi.org/10.1158/0008-5472.CAN-14-0796.
59. Wang, Y., Chen, Y.L., Huang, C.M., Chen, L.T., and Liao, L.D. (2023). Visible CCD Camera-Guided Photoacoustic Imaging System for Precise Navigation during Functional Rat Brain Imaging. Biosensors *13*, 107. https://doi.org/10.3390/bios13010107.
60. Tang, J., Xi, L., Zhou, J., Huang, H., Zhang, T., Carney, P.R., and Jiang, H. (2015). Noninvasive high-speed photoacoustic tomography of cerebral hemodynamics in awake-moving rats. J. Cereb. Blood Flow Metab. *35*, 1224–1232. https://doi.org/10.1038/jcbfm.2015.138.
61. Jo, J., Xu, G., Cao, M., Marquardt, A., Francis, S., Gandikota, G., and Wang, X. (2017). A Functional Study of Human Inflammatory Arthritis Using Photoacoustic Imaging. Sci. Rep. *7*, 15026. https://doi.org/10.1038/s41598-017-15147-5.
62. Xi, L., Jin, T., Zhou, J., Carney, P., and Jiang, H. (2017). Hybrid photoacoustic and electrophysiological recording of neurovascular communications in freely-moving rats. Neuroimage *161*, 232–240. https://doi.org/10.1016/j.neuroimage.2017.08.037.
63. Wang, Y., Chu, T.S., Lin, Y.R., Tsao, C.H., Tsai, C.H., Ger, T.R., Chen, L.T., Chang, W.S.W., and Liao, L.D. (2021). Assessment of Brain Functional Activity Using a Miniaturized Head-Mounted Scanning Photoacoustic Imaging System in Awake and Freely Moving Rats. Biosensors *11*, 429. https://doi.org/10.3390/bios11110429.
64. Wang, Y., Tsai, C.H., Chu, T.S., Hung, Y.T., Lee, M.Y., Chen, H.H., Chen, L.T., Ger, T.R., Wang, Y.H., Chiang, N.J., and Liao, L.D. (2022). Revisiting the cerebral hemodynamics of awake, freely moving rats with repeated ketamine self-administration using a miniature photoacoustic imaging system. Neurophotonics *9*, 045003. https://doi.org/10.1117/1.NPh.9.4.045003.
65. Sylolypavan, A., Sleeman, D., Wu, H., and Sim, M. (2023). The impact of inconsistent human annotations on AI driven clinical decision making. NPJ Digit. Med. *6*, 26. https://doi.org/10.1038/s41746-023-00773-3.
66. Van Rossum, G., and Drake, F.L. (1995). Python Reference Manual. (Centrum voor Wiskunde en Informatica Amsterdam).
67. Limited, R.C. (2024). PyQt5. https://riverbankcomputing.com/software/pyqt/intro.
68. Bradski, G. (2000). The OpenCV library. Dr. Dobb's J. *25*, 120.
69. van Dam, E.A., Noldus, L.P.J.J., and van Gerven, M.A.J. (2020). Deep learning improves automated rodent behavior recognition within a specific experimental setup. J. Neurosci. Methods *332*, 108536. https://doi.org/10.1016/j.jneumeth.2019.108536.
70. Kalueff, A.V., Aldridge, J.W., LaPorte, J.L., Murphy, D.L., and Tuohimaa, P. (2007). Analyzing grooming microstructure in neurobehavioral experiments. Nat. Protoc. *2*, 2538–2544. https://doi.org/10.1038/nprot.2007.367.
71. Berridge, K.C., Aldridge, J.W., Houchard, K.R., and Zhuang, X. (2005). Sequential super-stereotypy of an instinctive fixed action pattern in hyper-dopaminergic mutant mice: a model of obsessive compulsive disorder and Tourette's. BMC Biol. *3*, 4. https://doi.org/10.1186/1741-7007-3-4.
72. Kalueff, A.V., Stewart, A.M., Song, C., Berridge, K.C., Graybiel, A.M., and Fentress, J.C. (2016). Neurobiology of rodent self-grooming and its value for translational neuroscience. Nat. Rev. Neurosci. *17*, 45–59. https://doi.org/10.1038/nrn.2015.8.
73. Tzutalin (2015). LabelImg. https://github.com/tzutalin/labelImg.
74. Oliva, A., and Torralba, A. (2007). The role of context in object recognition. Trends Cogn. Sci. *11*, 520–527. https://doi.org/10.1016/j.tics.2007.09.009.
75. van den Buuse, M. (1999). Circadian rhythms of blood pressure and heart rate in conscious rats: effects of light cycle shift and timed feeding. Physiol. Behav. *68*, 9–15. https://doi.org/10.1016/s0031-9384(99)00148-1.
76. Salgado-Delgado, R., Angeles-Castellanos, M., Buijs, M.R., and Escobar, C. (2008). Internal desynchronization in a model of night-work by forced activity in rats. Neuroscience *154*, 922–931. https://doi.org/10.1016/j.neuroscience.2008.03.066.
77. Krinke, G.J. (2000). The Laboratory Rat (Elsevier).
78. Castelhano-Carlos, M., Costa, P.S., Russig, H., and Sousa, N. (2014). PhenoWorld: a new paradigm to screen rodent behavior. Transl. Psychiatry *4*, e399. https://doi.org/10.1038/tp.2014.40.
79. Blom, H.J.M., Tintelen, G., Baumans, V., Broek, J.D., and Beynen, A.C. (1995). Development and Application of a Preference Test System to Evaluate Housing Conditions for Laboratory Rats. Appl. Anim. Behav. Sci. *43*, 279–290. https://doi.org/10.1016/0168-1591(95)00561-6.
80. Stoynev, A.G., and Ikonomov, O.C. (1983). Effect of constant light and darkness on the circadian rhythms in rats: I. Food and water intake, urine output and electrolyte excretion. Acta Physiol. Pharmacol. Bulg. *9*, 58–64.
81. Willner, P., Towell, A., Sampson, D., Sophokleous, S., and Muscat, R. (1987). Reduction of sucrose preference by chronic unpredictable mild stress, and its restoration by a tricyclic antidepressant. Psychopharmacology (Berl) *93*, 358–364. https://doi.org/10.1007/BF00187257.
82. Ramirez, S., Liu, X., MacDonald, C.J., Moffa, A., Zhou, J., Redondo, R.L., and Tonegawa, S. (2015). Activating positive memory engrams suppresses depression-like behaviour. Nature *522*, 335–339. https://doi.org/10.1038/nature14514.
83. Cheeta, S., Ruigt, G., van Proosdij, J., and Willner, P. (1997). Changes in sleep architecture following chronic mild stress. Biol. Psychiatry *41*, 419–427. https://doi.org/10.1016/S0006-3223(96)00058-3.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: organisms/strains		
Rat: Sprague–Dawley	LASCO	N/A
Software and algorithms		
YOLOv5	open source website	https://github.com/ultralytics/yolov5
Python	Python software foundation	https://imagej.nih.gov/ij/
OpenCV	Python Data Analysis Library	http://opencv.org
PyQt	Riverbank Computing	https://www.riverbankcomputing.com/software/pyqt/
LabelImg	Tzutalin	https://github.com/HumanSignal/labelImg
Jetson Nano Developer Kit SD Card Image	NVIDIA Corporation	https://developer.nvidia.com/jetson-nano-sd-card-image
Microsoft® Excel® LTSC MSO (16.0.14332.20771)	Microsoft	Office 2021
Other		
Source code	This paper	https://github.com/hijay31/The_RW_system

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

All procedures were performed according to the Institutional Animal Care and Use Committee (IACUC) of the National Health Research Institutes (NHRI), Taiwan (IACUC number: NHRI-IACUC-111017). All animals were maintained in a housing environment with a constant temperature range of 23°C–25°C, a relative humidity range of 40–60%, a regular 12/12 h light/dark cycle, and *ad libitum* access to food and water. Standard rat cages suitable for housing one to two rats per cage were used in this study. When a single rat was housed in a single cage, two or more environmental enrichment items were provided to stimulate normal behavior and ensure the welfare of the animal. The environmental enrichment items used in this study were mainly nest materials and wooden strips. The training and verification sets included images of five healthy female Sprague–Dawley rats aged 5 to 16 weeks that were transferred from other experiments. Two animal experiments were conducted in this study: 1) a light cycle change experiment and 2) a sucrose preference test. The same 8 female Sprague–Dawley rats aged 16 to 17 weeks were used in both experiments, and the animals were sourced from BioLASCO Taiwan Co., Ltd., Taipei, Taiwan. One day before the experiment, each rat was individually housed, and a multifunctional camera was installed to allow the rats to acclimate to the environment.

METHOD DETAILS

Hardware of the multifunctional camera

The RW system hardware consisted of a multifunction camera and an edge computing device. Figure 1A presents an exploded view of the multifunction camera, showing all the components in detail, including an infrared lens (IC board model: LRCP5055W), six near-infrared lights, and a thermal lens (model: Lepton 3.5). The thermal lens has a longwave infrared range of 8–14 μm and a pixel size of 12 μm. The exterior and bracket of the camera were manufactured using a 3D printer (model: KINGSSEL K-3030). The multifunctional camera was installed in the home cage and fixed on the cage lid with a bracket and a screw at a 45-degree angle to monitor the rats. The wire was connected to a Jetson Nano edge device (model: NVIDIA Jetson Nano Developer Kit) outside the cage.

A near-infrared (NIR) lens with a field of view (FOV) > 160° was used for real-time identification of rat behavior. The complete FOV within the cage can be obtained, including images of the feed and water bottles located on the cage lid, as shown in Figure 1B. In experimental animal housing environments, a light cycle is generally implemented to simulate day and night. The NIR lens is set to display near-infrared images regardless of whether there is light in the cage and can be monitored throughout the day without being affected by changes in the light cycle. The thermal lens had an FOV of 57° and was used to track the rat's body temperature and location in real time. A photo of the actual installation of the camera in the housing cage is shown in Figure 1C. The camera was easy to install and fits in rat cages with standard sizes (26.0 × 47.6 × 20.3 cm, WxDxH). The modified cage lid reduced the problem of the feeder position blocking the view of the rat.

Program for determining location and body surface temperature

The RW system can be expanded for research applications and installed in multiple housing cages. The system automatically detects and records behavioral and physiological data of the rats. Researchers can monitor these data online in real time, helping reduce the burden associated with manual observation, as shown in Figure 2A. We used Python[66] and PyQt5[67] to develop a graphical user interface (GUI) software

tool for the RW system. This tool provides a convenient operation interface and real-time detection information, including behavioral recognition, location tracking, body temperature data, and pie charts of the proportion of each behavior. The body temperature detected by the RW system is the body surface temperature obtained through a thermal lens. The video and related data are stored in MP4 and CSV formats for further analysis, as shown in Figure 2B. The GUI runs on Windows and Linux. Figure 2C shows the program process of the RW system. When the multifunction camera is initiated, the NIR lens acquires the image. Then, an improved model based on YOLOv5 is implemented on the edge device, which can instantly identify a rat's behavior through deep learning technology and automatically display the bounding box of the area in which the recognized behavior was performed. Additionally, thermal lens images are used to obtain coordinates and pixel values to calculate the position and body temperature of the rat. Finally, when integrated into the GUI, the rat's behavior, body temperature, and position data can be displayed in real time, and historical behavior and body temperature information can be displayed using raster graphs, histograms, and curve graphs.

The rat cage was placed on a cage rack close to the wall, resulting in low light in the area near the cage wall. In addition, turning off the light at night resulted in no light in the cage; however, the surface temperature of the rat could still be detected with thermal imaging regardless of the brightness of the light. We used OpenCV[68] to control the temperature to a maximum of 42°C and a minimum of 29°C for the input temperature array. The aim of this step was to prevent interference from temperature values in the thermal images that clearly did not represent the rat's body temperature. Then, we performed image binarization, dilation, and erosion processing to obtain the center coordinates of the rat bounding box. A threshold for the pixel distance of the rat's movement was set to reduce the potential of unstable position changes, and unstable detection values higher than the threshold were filtered as outliers and not included in calculations. The 45-degree image was processed via checkerboard correction and perspective transformation to obtain a top-view image, which is better for visualizing the position of the rat.

The bounding box of the rat was obtained by processing the thermal image with OpenCV. Two modes were used for body temperature detection: in one mode, the center coordinate of the bounding box was used as the detection coordinate; in the other mode, the highest temperature value in the bounding box was used as the detection coordinate. The output temperature value was calculated from the moving average of the temperature values of the detected coordinates. A threshold for the moving distance of the coordinates was also set. The detected coordinate values higher than the threshold were filtered as outliers, and a new coordinate was obtained from the next frame for use in calculations.

Animal behaviors are complex and diverse. We defined eight behavioral categories for this study with reference to common behavioral classifications in behavioral research. The categories included drinking, eating, grooming, rearing, gnawing, searching, resting, and scratching. Sniffing behavior is a major activity in the daily lives of rodents.[69] We classified sniffing behavior in the searching category and defined searching as the default behavior, which included unclassified behaviors and partial images with limited viewing angles. The behavior classification diagram is shown in Figure 3. Multiple behavioral postures were classified as grooming. Rodent self-grooming behavior has a complex and sequential structure consisting of repetitive stereotyped behaviors. The grooming behavior includes paw paddling around the nose in the first phase, facial grooming in the second phase, bilateral paddling backward and upward with both paws in the third phase, and body grooming in the fourth phase.[70,71] Furthermore, tail and genital grooming can often be observed in rodents as part of their general grooming behavior.[72]

Dataset structure and sources

The training set consisted of images of one to two rats per cage collected from the cage racks of the housing room between May 2022 and December 2022. A multifunctional camera installed in the home cage collected a total of 34 videos, each with a duration of 1 h. In each video, 1–2 images were captured per second, and the resolution of each image was adjusted to 640 × 480 pixels. The camera was installed one day before video collection to allow the animals to acclimate to the presence of the camera. There were 14,123 images in the training set with 15,158 labeled instances, which were selected and labeled by an experienced researcher using the LabelImg[73] tool. The number of instances of each behavior category is shown in Figure 4A. The proportion of rat behaviors in the searching category was relatively high, so the researchers selected more images of searching behaviors to include in the training set. The training set contained 11,189 images of a single rat and 1,987 images of 2 rats, as shown in Figure 4B. In addition, 6.7% of the images in the training set were animal-free background images, which were included to help reduce the number of false positives during model recognition. The verification set was labeled by the same researcher who labeled the training set and was used to adjust the model training parameters and optimization function to find an optimal model. The verification set images were obtained from two 1-h video files captured in February 2023. The rat captured in these videos was one of the five rats from the videos in the training set; however, the recording period did not overlap with that for the training set. The two videos contained 7,200 images in total; in each video, one image was captured per second, resulting in 3,600 images per video, as shown in Figure 4C. The images in the validation set were captured from consecutive video frames and represent the distribution of the rat's behavior over an hour. The analysis revealed that the rat performed searching and resting behaviors more than the other behaviors, as shown in Figure 4D.

The test set plays a crucial role in verifying the generalizability of the model and ensuring its effectiveness in real-world applications. The four subsets of the test set included consecutive video frames from four videos taken of four rats the day before the light cycle change experiment. The four subsets were labeled test sets 1, 2, 3, and 4, and each subset contained 3,600 images. Analysis of the test sets revealed that, except for test set 1, the number of images in the searching and resting categories was greater than the number of images in the other categories. In contrast, the number of images in the resting category was higher in test sets 2, 3, and 4, with fewer instances of this category appearing in test set 1, as shown in Figure 4E.

Model training

The YOLO series algorithms are popular real-time object detection algorithms and are widely used in computer vision. The proposed model was developed based on the YOLOv5s neural network architecture. The model has 7.2 million parameters and a file size of only 14 MB, making it suitable for running on edge computing devices. The hardware for model training included an i7-13700K processor, 64 GB of RAM, and an NVIDIA GeForce RTX 4090 24 GB GPU. The training hyperparameters were set to a batch size of 32, momentum of 0.937, weight decay of 0.0005, and learning rate of 0.01. The YOLOv5s pretrained model was used for training, and early stopping was performed during training. The training dataset was used to train the model, whereas the validation dataset was used to evaluate the model's performance and generalizability during training. After each epoch, inference was performed based on the validation dataset, and the corresponding loss and evaluation metrics were calculated. By monitoring the changes in these metrics, the convergence of the model during training can be evaluated.

Contextual object labeling

Contextual object labeling is a method in which the surrounding environment of an object is considered in object detection or object recognition tasks. In the real world, target objects often do not exist in isolation; they are interconnected with other objects and specific environments. This provides rich background information that can be used by the visual system to aid in object recognition.[74] Because of the type of cage we used, the water bottle was placed inside the feeder, as the rat's mouth needed to touch the outlet of the water bottle for the rat to drink. There was an obvious relationship between the rats' drinking behavior and the water bottle cap. This allowed us to include the stainless steel water bottle cap within the bounding box of the rat's drinking behavior, thus enhancing the contextual information associated with the drinking behavior.

Modification of nonmaximum suppression

Nonmaximum suppression (NMS) is an important postprocessing step in object detection. Deep learning CNNs have been applied in target detection tasks; neural network models are usually used to detect and output sets of instances. Each instance includes the bounding box location, class prediction, and confidence score. Overlapping instances can be removed using NMS, with only the final instances output. In the focus category of this study, the postures of some resting and searching behaviors are similar, which affects the confidence score and leads to incorrect identification in the final NMS output. We modified the confidence levels of the resting and searching categories for a set of instances detected by the model. Then, we removed overlapping instances via NMS. After NMS processing, the instance with the highest confidence score was used as the final output result.

The NMS method was modified as follows. At the beginning of the model detection process, the Euclidean distance E between the detected bounding boxes of the previous frame b_{i-1} and the prior previous frame b_{i-2} was calculated. If the distance was lower than the set threshold D_t, the object was determined to be in a stationary state. However, if the distance was greater than the threshold D_t, the object was determined to be in a nonstationary state. The status of the next frame was repeatedly determined, and the number of consecutive stationary states S was calculated, as shown in pseudocode.

$$S\begin{cases} S+1, & E(b_{i-2}, b_{i-1}) \le D_t \\ S=0, & E(b_{i-2}, b_{i-1}) > D_t \end{cases} \quad \text{(Equation 1)}$$

If the number of consecutive frames in the stationary state S was greater than the set threshold N, the object was determined to be in a stationary state. Within the set of instances s_i detected by the model, all instances in the resting category $s_i(rest)$ were multiplied by x_{up} to increase their confidence scores. Additionally, all instances in the searching category $s_i(search)$ were multiplied by x_{down} to reduce their confidence scores. Conversely, if the object was determined to be in a nonstationary state, the confidence scores of all instances in the resting category were decreased, and the confidence scores of all instances in the searching category were increased, as shown in pseudocode.

$$s_i\begin{cases} s_i(rest)x_{up}; s_i(search)x_{down}, & S > N \\ s_i(rest)y_{down}; s_i(search)y_{up}, & S \le N \end{cases} \quad \text{(Equation 2)}$$

Evaluation metrics

We used evaluation metrics that are commonly applied in deep learning. The precision represents the accuracy of the model in predicting positive examples. The recall represents the model's ability to identify positive examples. The number of true positives (TPs) indicates the number of times that positive categories were correctly predicted by the model as positive, whereas the number of false positives (FPs) represents the number of times that positive categories were incorrectly predicted. The number of true negatives (TNs) is the number of times that negative categories were correctly predicted, and the number of false negatives (FNs) is the number of times that negative categories were incorrectly predicted. The calculation formulas are shown as follows:

$$Precision = \frac{TP}{TP+FP} \quad \text{(Equation 3)}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{(Equation 4)}$$

The precision−recall (PR) curve is a graphical representation used to evaluate the relationship between the precision and recall of a binary classification model at different thresholds. Ideally, the PR curve should be closer to the upper right corner of the coordinate axis, indicating that the model can maintain both high precision and high recall. This enables the model to capture more positive examples successfully while reducing the number of FP predictions.

The average precision (AP) is a metric commonly used in information retrieval and object detection tasks. It is calculated by computing the PR curve for a model and then calculating the area under that curve. In other words, the AP represents the tradeoff between precision and recall across different thresholds. The calculation formula is shown as follows:

$$AP = \int P(R)dR \quad \text{(Equation 5)}$$

The F1 score is a metric used to evaluate the performance of a binary classification model. The precision and recall are both considered to calculate a single score ranging from 0 to 1, with higher values indicating better model performance. The calculation formula is shown as follows:

$$F1\ \text{score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad \text{(Equation 6)}$$

The intersection over union (IoU) metric is commonly used in computer vision and image analysis. It is employed to evaluate the accuracy of object detection algorithms, particularly in tasks such as object localization and instance segmentation. The IoU quantifies the degree of overlap between a predicted bounding box B_p and an actual bounding box B_g. The calculation involves determining the ratio of the intersection area between the predicted bounding box B_p and the actual bounding box B_g to the union area. This ratio is then used to assess the detection effectiveness, distinguishing between TPs and FPs. The calculation formula is shown as follows:

$$IoU = \frac{|B_p \cap B_g|}{|B_p \cup B_g|} \quad \text{(Equation 7)}$$

Animal experimental research design

Rodents are primarily nocturnally active animals,[75–78] and albino rats prefer low light intensity in housing cages.[79] The light cycle for experimental animals typically consists of 12 h with lights on and 12 h with lights off. Photoperiod changes can affect animal physiology and behavior.[80] We conducted a light cycle change experiment for two consecutive days and recorded the changes in animal behavior every second. On the first day of the experiment, eight rats in the light/dark (L/D) group were subjected to a normal 12-h light/12-h dark cycle. The next day, the light cycle was changed to 24 h of continuous light, resulting in the light/light (L/L) group. Each group of rats was divided into two batches. After the first batch of 4 rats was subjected to the experiment with 4 sets of multifunctional cameras, the second batch of 4 rats was subjected to the same experiment. Due to abnormal file reading, 14 h of data for rat number 1 on the first day could not be obtained and was thus excluded from the data analysis. The experimental dataset included data from seven rats, with 51 min of data unavailable for rat number 4 on the first day because of abnormal file reading. In the data analysis, drinking, eating, grooming, rearing, gnawing, and scratching were defined as activity.

The sucrose preference test (SPT) is a behavioral experiment method typically used to study emotions, stress, and depression in animals. The test reflects an animal's emotional state and depression type through its preference for sweet substances.[49,81–83] In a typical SPT, animals are exposed to two water bottles. One bottle of water does not contain sucrose, while the other does, and the animals can choose freely between the two bottles. Finally, the intake from each bottle is measured to calculate the preference for sucrose water. This experiment allows researchers to assess how animals respond to pleasure and reward. The percentage concentration of sucrose in the water bottles used in this experiment was 5%. Eight rats were given two bottles of sucrose-free water on the first day, referred to as the water/water (W/W) group, and two bottles of sucrose-containing water on the second day, referred to as the sucrose/sucrose (S/S) group. On the third day, the rats were given a sucrose-free water bottle and a sucrose-containing water bottle, referred to as the water/sucrose (W/S) group, to test their preferences. Each group of rats was divided into two batches. After the first batch of 4 rats was subjected to the experiment with 4 sets of multifunctional cameras, the second batch of 4 rats was subjected to the same experiment. The sucrose preference was calculated as 100% × time spent drinking (s)/total experimental time (s). During the experiment, the rats had *ad libitum* access to food and water. The experimental dataset included data from eight rats, with data from rats 2 and 3 in the W/S group unavailable for 44 and 166 min, respectively, on the third day due to abnormal file readings.

In this experiment, the RW system was used to identify drinking behavior, track the location of the rat, and detect the body temperature of the rat. We divided the housing cages into four areas. The coordinates of the rats recognized by the RW system corresponded to these areas. The first and second areas corresponded to the positions of the two water bottles in the housing cage. The RW system determined the time a rat spent drinking from the two water bottles by analyzing the locations and drinking behaviors of the rats, after which the preferred water bottle was determined. The rat localization data in this experiment were obtained via model recognition rather than thermal data because the model was updated after the experiment.

QUANTIFICATION AND STATISTICAL ANALYSIS

All the values are expressed as the mean ± SEM. Paired two-sided t-tests were performed for the light cycle change test and sucrose preference test using Excel (Microsoft Software). The statistical significance of the differences in the recognition accuracies of the models (O, D, and RD) was assessed via McNemar's test. The results of each model were evaluated against the ground truth labels of the test set, resulting in binary classification outcomes for each instance. Pairwise comparisons were performed with McNemar's test, a nonparametric method for evaluating differences in paired proportions. Specifically, McNemar's test was used to determine whether the differences in the prediction accuracies of the O and D models and the D and RD models were statistically significant. For each comparison, a 2x2 contingency table was constructed, with cells representing counts of TPs, FPs, FNs, and TNs. The test was conducted with 1 degree of freedom. Statistical significance was determined at thresholds of $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$.