Descriptor

Patterns

Highly accurate and precise determination of mouse mass using computer vision

Highlights

- Method to non-invasively determine mouse body mass using video data
- Predicts mass with ~5% mean error across sexes and strains with high visual diversity
- Enables continuous monitoring of mouse mass without human intervention
- Visual mass assessment extends to multiple mice in home environments

Authors

Malachy Guzman, Brian Q. Geuther, Gautam S. Sabnis, Vivek Kumar

Correspondence

vivek.kumar@jax.org

In brief

Guzman et al. develop computer vision models to non-invasively determine the body mass of a mouse from video data. Accurate tracking of body mass is an important aspect of many mouse-model experiments and of ethical care standards. This approach allows continuous tracking of mice in complex environments, which improves experiment and care quality without the confounding physiological effects of manual handling.

Guzman et al., 2024, Patterns *5*, 101039
September 13, 2024
https://doi.org/10.1016/j.patter.2024.101039

CellPress

Patterns

CellPress
OPEN ACCESS

Descriptor

Highly accurate and precise determination of mouse mass using computer vision

Malachy Guzman,[1,2] Brian Q. Geuther,[1] Gautam S. Sabnis,[1] and Vivek Kumar[1,3,4,5,*]
[1]The Jackson Laboratory, Bar Harbor, ME, USA
[2]Carleton College, Northfield, MN, USA
[3]School of Graduate Biomedical Sciences, Tufts University School of Medicine, Boston, MA, USA
[4]Graduate School of Biomedical Sciences and Engineering, University of Maine, Orono, ME, USA
[5]Lead contact
*Correspondence: vivek.kumar@jax.org
https://doi.org/10.1016/j.patter.2024.101039

THE BIGGER PICTURE The laboratory mouse is the primary tool to model human diseases, study mechanisms, and carry out preclinical efficacy and toxicity testing of novel therapeutics. Animal experiments often rely on body mass as a key indicator of health. However, the frequent handling by the tester that is required to weigh animals on a scale can induce physiological stress responses and potentially confound the experimental results. Here, we used machine vision—the combined application of machine learning and computer vision—to non-invasively determine the body mass of a mouse. Our methods allow for mass tracking in visually complex environments without human interference. In biomedical contexts, our tool improves experiment and husbandry quality and ethical care standards through continuous acquisition of accurate, precise, and autonomous mass data.

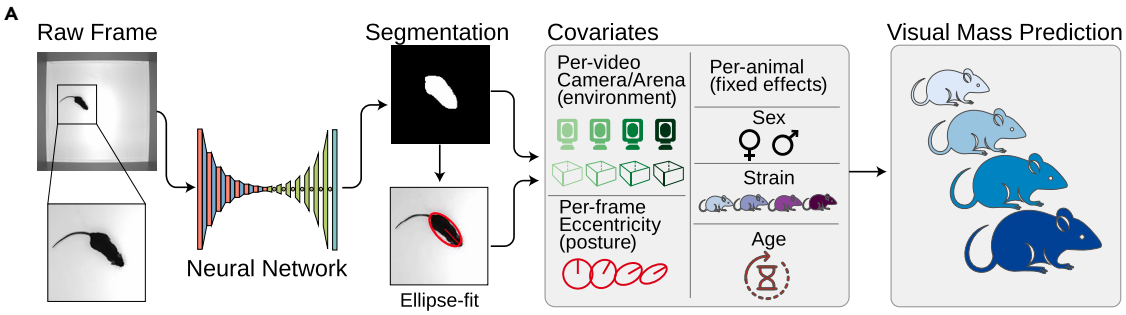
SUMMARY

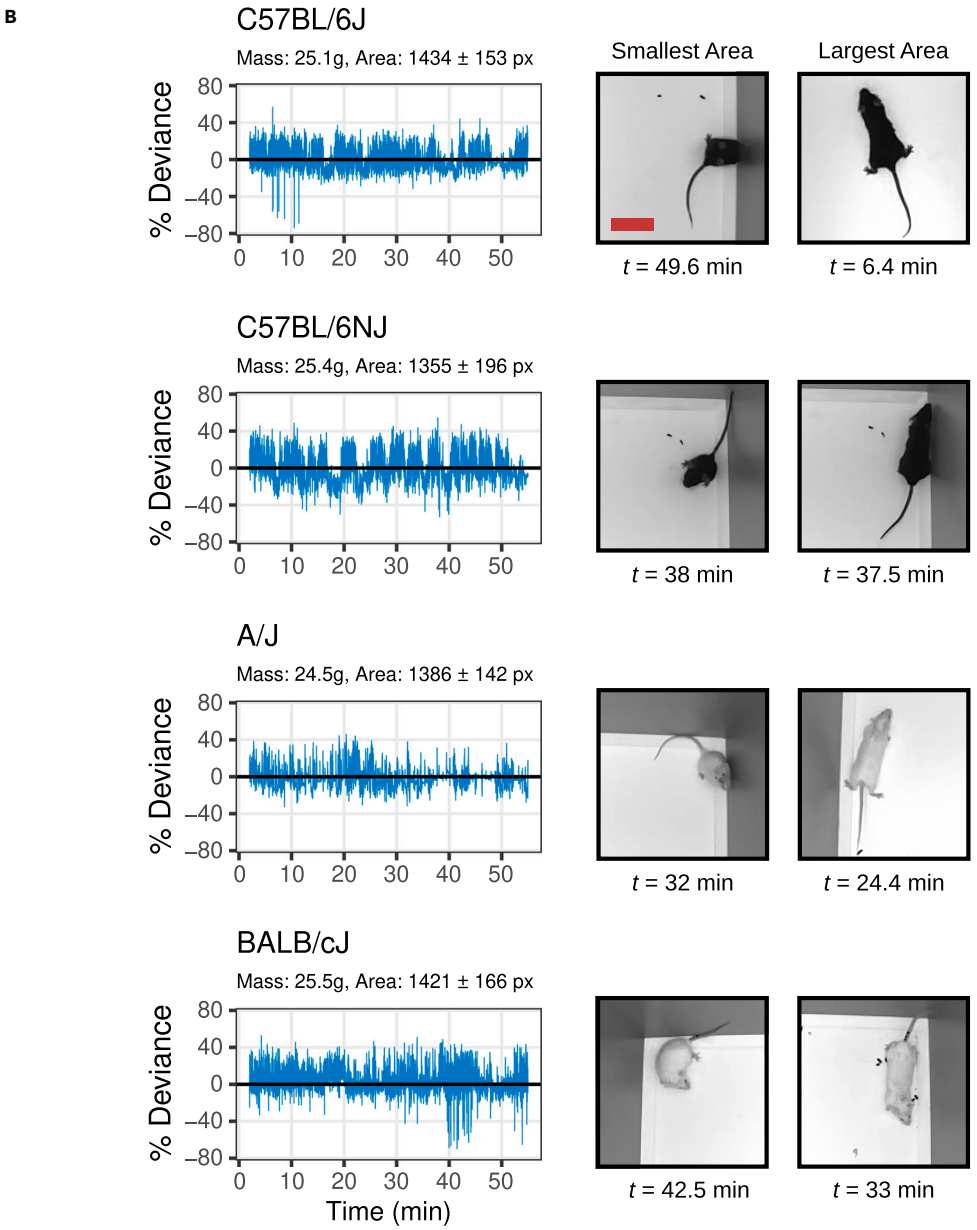
Changes in body mass are key indicators of health in humans and animals and are routinely monitored in animal husbandry and preclinical studies. In rodent studies, the current method of manually weighing the animal on a balance causes at least two issues. First, directly handling the animal induces stress, possibly confounding studies. Second, these data are static, limiting continuous assessment and obscuring rapid changes. A non-invasive, continuous method of monitoring animal mass would have utility in multiple biomedical research areas. We combine computer vision with statistical modeling to demonstrate the feasibility of determining mouse body mass by using video data. Our methods determine mass with a 4.8% error across genetically diverse mouse strains with varied coat colors and masses. This error is low enough to replace manual weighing in most mouse studies. We conclude that visually determining rodent mass enables non-invasive, continuous monitoring, improving preclinical studies and animal welfare.

INTRODUCTION

Body mass is a primary measure of health and disease in humans. For example, the body mass index is a key measure of metabolic health and is one of the oldest and most widely used metrics in modern medicine.[1] Changes in body mass indicate the function of many systems, including metabolic, cardiac, and psychiatric, and are often the primary symptom of disease onset.[2,3] Rodents, particularly mice, are commonly used to model human diseases, carry out preclinical studies, and investigate disease pathologies. Over 95% of disease research that involves animal models is conducted with mice.[4] As in humans, changes in body mass in mice are predictive of health, particularly when sudden changes occur.[5] The Institutional Animal Care and Use Committee (IACUC) includes significant body mass loss in its "humane intervention" guide, a set of standardized criteria that call for veterinary intervention or euthanasia of subjects. Loss of body mass greater than 20% relative to baseline or matched controls is a common justification for such intervention.[6]

Outside of ethical care, body mass is also an important feature collected during preclinical, metabolic, cardiovascular, and neurobiological studies.[7] However, handling mice to assess mass induces physiological stress responses and alters immune responses.[8,9] Additionally, weight measurement is static; body mass is typically measured every few days during the animal's rest phase, depending on the protocol. Weight is often visually assessed during daily health checks by caretakers, and mass





(legend on next page)

is manually measured only if the animal looks unhealthy or has lost weight. This can be subjective and can lead to variable treatment of the animal. These effects of human handling and static periodic measures of mass can weaken the validity and reproducibility of mouse experiments. Thus, there is a need for better methods to accurately and non-invasively measure animal mass over time in order to improve mouse models of disease and improve animal welfare.

To address this problem, we explore a generalized solution using computer vision (CV) to determine body mass. The application of computer vision to determining an animal's body mass has been performed in industrial farming, including poultry farming, cattle, and fishing.[10–13] Additionally, a recent approach to automating a frailty index for mice uses CV-based segmentation masks to predict body mass as a potential measure of frailty, achieving a similar R^2 value to our basic M1 model described below, though with higher error.[14] Existing CV approaches for determining mass in farming contexts use a multitude of visual metrics, including top-down silhouette area, eccentricity, perimeter, body length, and body width—similar metrics to the ones we consider. These solutions use a wide range of modeling approaches, typically some form of linear regression or some type of neural network applied to predict mass directly from the image.[12] Depth cameras have also been utilized for a better 3D representation of the animal.[15–18] The central difference between our solution and most industrial farming approaches is that errors introduced by animal posture are largely an issue of the animal not standing in an optimal location.[10] Visual mass determination in these contexts is thus an easier task, as cattle tend to have rigid postures and relatively constant silhouettes compared to mice. Several studies dealing with visual mass prediction for cattle often achieve R^2 values above 0.9 and relative error under 10%.[10,19] Other farming applications deal with the prediction of carcass mass, where posture is even less of a factor.[11] Mice, on the other hand, are highly flexible, have much smaller but highly deformable bodies, and quickly change shape. This deformability is a direct result of animal behavior and is subject to individual differences in genetics. Thus, the deformability of rodents is a major challenge to visual weight assessment, which we approach by combining two visual metrics, area and eccentricity, and applying multiple linear regression.

Other non-invasive methods to determine animal mass have been implemented using highly engineered cages with balances and compartments.[5] Although these methods alleviate animal handling issues and provide continuous mass measurement, the highly engineered cage designs require modification of housing conditions and are currently limited to singly housed animals. Thus, these methods can be challenging to scale and practically implement. A computer vision approach is suitable for multiple environments, both within and outside home cages, and can be scaled for continuous monitoring of mass in multiple mice. Therefore, this approach is potentially highly generalizable and scalable. However, whether this approach is accurate and precise enough to replace manual weight assessment, particularly in genetically diverse mouse strains, has yet to be determined.

Here, we test the use of single-camera video surveillance that functions autonomously and does not require the animal to take predetermined actions. We compare 6 statistical models with different use cases on a dataset of 62 genetically diverse mouse strains and show highly accurate and precise visual mass assessment. Our approach is applicable to preclinical studies, mouse husbandry, and other biological studies. It has the potential to expand weight assessment to novel environments and for continuous assessment. Additionally, this method can be easily incorporated into existing computer-vision-based monitoring systems for routine use. We conclude that computer-vision-based mass determination is highly accurate and precise and can be further developed for real-world applications.

RESULTS

General approach

Our approach to visual mass determination is outlined in Figure 1A and detailed below. Briefly, we collect video data from the top-down perspective in an open field.[20] Each frame is segmented for the mouse, which describes the animal's size. The segmented image is used to fit an ellipse that describes the approximate posture of the animal and is then adjusted with covariates to improve our modeling and mass prediction.

Mice are highly deformable, depending on their behavior. To determine the effect of deformability on the segmentation area, we inspected the variation of segmentation area in 55-min videos of several mice that are approximately the same mass (25.1 g) from different strains (Figure 1B). We found that the area of the mouse commonly deviated by ±40% (relative to the mean area) over the course of the video (Figure 1B, left). Upon examination of the video frames, we saw that the high variation in area was due to deformability of the animal due to altered behavior (Figure 1B, right). Over a short time frame, the mouse scrunches, stretches, bends, and rears such that its segmentation area quickly changes.

Since we have multiple images for each animal, we focus on the mean or median segmentation area. However, variance in this measure can impact the precision of a prediction. We investigated this by exploring changes in the variation of segmentation area between these mice. We found that this variation is strain specific, i.e., dependent on genetics. For instance, a 25.4-g C57BL/6NJ mouse has an average area of 1,355 ± 196 px, whereas a smaller, 24.5-g A/J mouse has a larger area of 1,386 ± 142 px. However, the variation in area is larger in C57BL/6NJ than in A/J mice. A/J mice are a high anxiety strain with low ambulation, while C57BL/6NJ mice have more bouts of activity.[21,22] The larger segmentation area and lower variation in A/J mice is because they spend more time in the corner in a constant posture. We reasoned that simple segmentation is inadequate to handle the genetic diversity seen in laboratory mice, and a more sophisticated approach is needed.

Figure 1. Visual mass determination approach taken to address highly variable segmentation areas observed in our data

(A) Flow chart describing the full computational process from open-field video to body mass prediction.

(B) Time series of percentage of deviations from the mean segmentation area over 55 min for four individual C57BL/6J, C57BL/6NJ, A/J, and BALB/cJ mice with (mean ± SD) of pixel area reported. Raw frames of the approximate least and greatest segmentation areas are shown to the right. Red bar on the C57BL/6J smallest frame indicates 5 cm.

Table 1. Posture adjustment metrics

Model	Input variable	Metric Formula	RMSE (g)	MAE (g)	MAPE (%)	R^2
T1	A_{cm} × *Med(eccentricity)*	$(w^2+l^2)^{1/2}/l$	2.353 ± 0.054	1.759 ± 0.038	7.328 ± 0.173	0.827 ± 0.01
T2	A_{cm} × *Med(aspect ratio)*	w/l	3.569 ± 0.081	2.830 ± 0.06	12.298 ± 0.275	0.603 ± 0.022
T3	A_{cm} × *Med(elongation)*	$\frac{\mu_{20}+\mu_{02}+\sqrt{4(\mu_{11})^2+(\mu_{20}-\mu_{02})^2}}{\mu_{20}+\mu_{02}-\sqrt{4(\mu_{11})^2+(\mu_{20}-\mu_{02})^2}}$	16.83 ± 15.77	4.714 ± 0.414	21.059 ± 2.196	0.170 ± 0.125
T4	A_{cm} × *Med(speed)*	$\sqrt{(v_x)^2+(v_y)^2}$ or $\lvert\vec{v}\rvert$	5.644 ± 0.127	4.422 ± 0.095	19.487 ± 0.460	0.002 ± 0.003

Models T1–T4 each have one input variable, the product of A_{cm}, and the median of the given metric. T1's input is equivalent to A_e, and the inputs of T2–T4 are analogous to A_e but for aspect ratio, elongation, and speed. In the metric formulas, w and l are the width and length of the fitted ellipse, v_x and v_y are the x and y components of velocity (also representable as $\vec{v}$), and each μ_{ij} is a central image moment, calculated with the python package OpenCV.

Data and data acquisition

Our dataset consists of 2,028 videos obtained from a previously conducted strain survey experiment.[20] For an example of our video data, see Video S1, where we provide 10-s clips from the videos corresponding to the frames in Figure 1B. Sampled mice come from 62 different strains, with most strains represented by at least 4 male and 4 female mice (Table S1). Our sampled mice range in age from 6 to 81 weeks old, with a mean age of 11 weeks. Body mass was measured immediately before open-field recording with a precision laboratory scale and ranges from 9.1 to 54 g with a mean of 24.7 g (see Figure S1 for age and mass distributions). Each video has one mouse in an open-field enclosure, which measures 52 × 52 cm. Videos are 480 × 480 px, were recorded at 30 fps, last 55 min, and have 8-bit monochrome depth. The camera is positioned approximately 36 inches (91.44 cm) above the floor. A 25.1-g C57BL/6J mouse occupies 1,434 px on average, or 0.6% of the total image pixels (Figures 1A and 1B).[23] This is similar to the distance and mouse pixel area if a camera were placed on top of a mouse cage changing table used in most barrier mouse facilities.

Error metrics

To evaluate fit accuracy, we used the coefficient of determination R^2. To evaluate model error, we used mean absolute error (MAE), root-mean-square error (RMSE), and mean absolute percentage error (MAPE) (see the experimental procedures for detail).[24–26] We used MAE to get an intuitive sense of average error, RMSE to penalize larger errors more strictly, and MAPE to get error relative to mass. Each of these metrics is commonly used to evaluate the performance of similar computer vision solutions.

Per-frame segmentation

First, we process a raw video from the dataset through one deep neural network to predict a segmentation mask for the mouse for every frame of the video (Figure 1A). The segmentation network has been trained on a diversity of mouse images and achieves high accuracy.[20] We fit an ellipse to the segmented blob as an approximation of mouse posture.

In every video, we compute the pixel area of the segmented mouse image in each frame by summing the number of "mouse" pixels, which are identified by the segmentation mask. The white pixels in the "segmentation" frame of Figure 1A are an example of mouse pixels. We summarize each video by the median of the pixel areas across all frames in the video, denoted A_{px}. This provides a single area measurement per mouse per video that we can use for modeling.

Per-camera and -arena correction

We collect data in 24 open-field arenas. Although we have standardized video data collection to a tight tolerance, there is still slight variation in image acquisition between arenas. To account for differences in camera lens zoom, we converted the measured segmentation (square pixel units) of the videos to real-world metric units (square cm units). Since all our arenas are manufactured to measure 52 × 52 cm, we located each arena's corners in each video by applying a corner detection deep neural network, thereby deriving a scaling factor from pixels to centimeters.[27] To account for arena variance, we adjust the segmentation measurement A_{px} by the following equation: $A_{cm} = A_{px}/(L_{px}/52\ \text{cm})^2$, where A_{cm} is the area in square cm, A_{px} is the area in square pixels, and L_{px} is the corner edge length in pixels. This adjustment does not necessarily require knowledge of in which arena the mouse was measured; one only needs to know the real-world distance between corners, which should be a constant between all arenas in an experiment. If an experiment uses arenas of intentionally varying size, then A_{cm}'s equation would have an arena-based variable instead of 52 cm.

Posture correction

So far, we have identified a metric, A_{cm}, that captures the segmentation area of a single mouse, normalized for differences in camera zoom. However, this metric does not account for the wide variation in area for individual mice over short time periods observed in Figure 1B, which we determined was a direct result of changing behavior and posture in the mice. We reasoned that accounting for varying posture is necessary to increase the performance of our visual method.

To correct for posture, we examined the effect of normalizing the area based on geometric shape descriptors of the segmentation mask. We tested several geometric shape descriptors, including eccentricity, aspect ratio, and elongation, as well as the speed of the mouse, as defined and evaluated in Table 1. We tested these shape descriptors under the assumption that changes in area in a short time frame are not related to changes in mass and so must be related to changes in pose. To determine which shape descriptor works best, we compared RMSE, MAE, MAPE, and R^2 values for each of four models (T1–T4, Table 1).

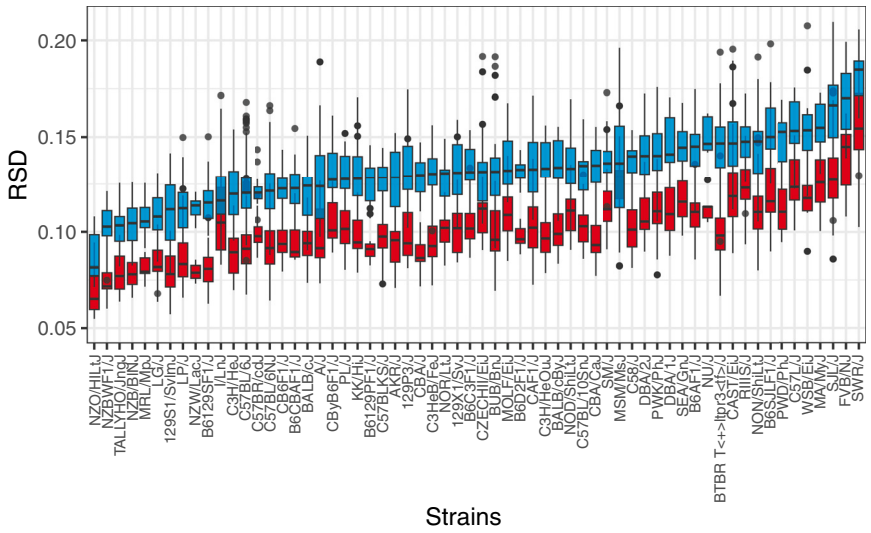


Figure 2. Adjustment of segmentation area by eccentricity reduces RSD across strains

Relative standard deviations (RSDs) of segmentation area A_{cm} (in blue) and the adjusted size A_e (in red) for each of the 62 mouse strains considered in our dataset. Each box shows the interquartile range (IQR) and the black whiskers reach the farthest point within 1.5 IQR of the end of the box.

Each model is a single-variable linear regression, taking the product of A_{cm} and the median of one posture metric (eccentricity, aspect ratio, elongation, or speed) as its input variable and predicting body mass. These models are evaluated under 50-fold cross-validation with a 70–30 training-testing split, explained in full detail in the statistical analysis.

Model T1 uses hyperbolic eccentricity—defined as $e = (w^2+l^2)^{1/2}/l$, where w and l are width and length, respectively—as its shape descriptor, T2 uses aspect ratio, T3 uses elongation, and T4 uses speed, all defined in Table 1. With hyperbolic eccentricity, T1 achieves an R^2 of 0.827 and a MAPE of 7.328%, significantly better performance than models T2–T4. This led us to conclude that eccentricity is the most useful indicator of shape of the metrics we tested. We initially hypothesized that speed would be a good descriptor by which to correct posture since mice tend to take on a constant elliptic shape when they are moving. However, speed performs poorly, resulting in a MAPE of 19.4% and an R^2 value of only 0.002. We hypothesize that this lack of quality is due to the proportion of time mice actually move quickly. Although mice do take on relatively constant postures while walking forward, this kind of motion takes up a relatively tiny fraction of total frames and is likely strain specific. Thus, speed may work for certain strains but does not generalize well in genetically diverse populations.

From this result, we formalize the "eccentric area" metric as $A_e = A_{cm} * e$, where A_{cm} is the median of the unit-converted area and e is the median eccentricity, both over all frames of the given video. In full, $A_e = (A_{px}/(L_{px}/52cm)^2) * (w^2+l^2)^{1/2}/l$. A_e functions as our video-level summary metric. To quantify the effect of correcting posture with eccentricity, we compared the variation in A_{cm} (anecdotally shown to be high in Figure 1B) and A_e by computing their respective relative standard deviations (RSDs) across all strains (Figure 2, A_{cm} in blue vs. A_e in red). Note that these RSDs are of the observed area values and not necessarily comparable to the RSDs of model prediction (see the supplemental information for more detail).

Strain is another important factor in variability, as some strains, like the obesity model NZO/HILtJ, have a low RSD of about 6%, whereas a wild-derived strain like WSB/EiJ has an RSD of about 15%. The largest area RSD we observed was about 18%, for the SWR/J strain. We hypothesize that these strain-level differences in RSDs are largely due to differing activity levels and behavioral patterns between strains. After posture correction, the RSD across all strains is significantly reduced (Figure 2, in red). The RSD, however, is still strain dependent, indicating that there are still strain-level effects on the precision of our prediction even after posture correction.

Other covariate correction

Equipped with several ways to describe size (A_{px}, A_{cm}, and A_e), we built six multiple linear regression models (M1–M6) to predict body mass. Each model uses raw area (A_{px}), unit-converted area (A_{cm}), or eccentric area (A_e) as its visual variable and different subsets of sex, strain, age, and arena (the specific open field the mouse was measured in) as covariates (Figures 3A–3D). After building each model, we performed a 50-fold cross-validation on a 70/30 training/testing split. Averaging our reported accuracy and error values over these 50 iterations ensures that our analysis is not biased by lucky or unlucky sampling.

Our first model, base (M1), is a single-variable linear regression that uses the raw segmentation area A_{px} to predict body mass, which we used as a comparative baseline for the other models (Table 2). This model has an R^2 value of 0.767 and a MAPE of 8.584% (Figure 3A). The second model, M2, is also a single-variable regression but between the unit-converted area A_{cm} and body mass. M2 performs better than the base model, increasing accuracy and decreasing error. In M3, we added a second variable, arena, which is the particular arena in which a mouse was tested. M3 performs slightly better than M2 by all measures. M4 introduces eccentricity, which is combined with the unit-converted area to be one variable, A_e, as previously described; M4 has two variables: A_e and arena. M4 performs substantially better, with an R^2 of 0.822 and a MAPE of 7.566%.

In a practical case of assuming minimal information, one would probably use M4 instead of M1, M2, or M3 because A_{px}, A_{cm}, and A_e are visual metrics, and the particular arena is inherent to the experiment. We could always measure these variables and would have no reason not to; thus, not using M4, in this case, would decrease prediction performance for no experimental benefit. However, we present each model here because selectively adding and/or modifying variables demonstrates the performance benefit of each unit conversion, arena

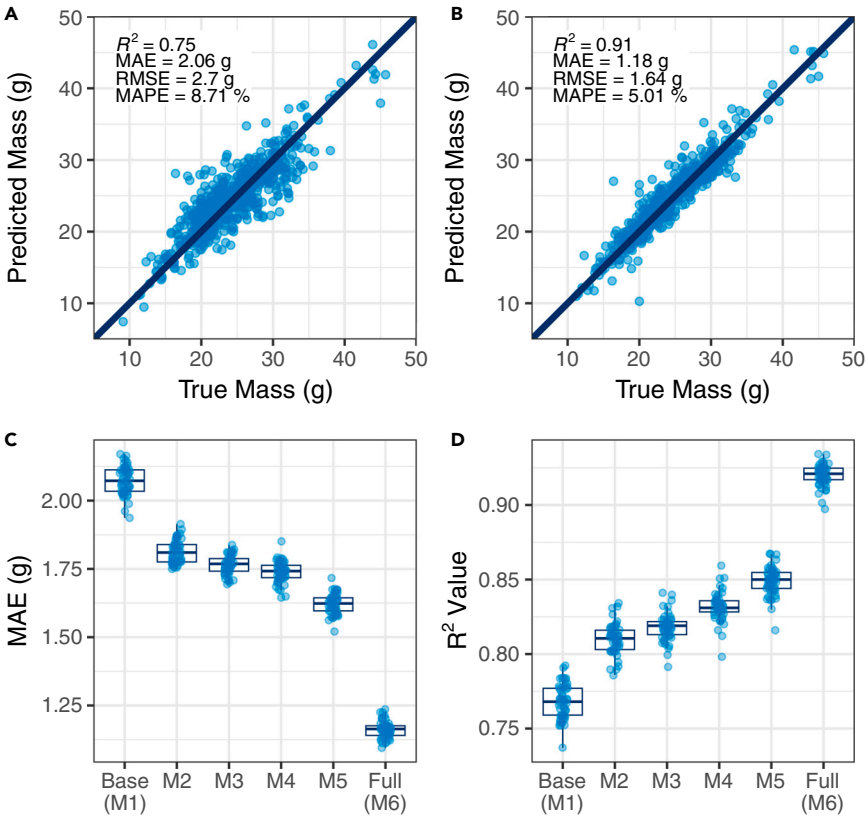


Figure 3. Model performance

(A and B) Comparison of true mass, as measured with a regular scale, and predicted mass, as determined by our process. Data shown are a one-test sample. R^2, mean absolute error (MAE), root-mean-square error (RMSE), and mean absolute percentage error (MAPE) are presented. The 45° line indicates a perfect prediction.

(A) Prediction quality for the base model (M1) accounting for only segmentation area.

(B) Prediction quality for the full model (M6) accounting for true size A_e, arena, sex, age, and strain.

(C and D) Error and accuracy performance for each model under 50-fold cross-validation.

(C) MAE across models. Each box shows the interquartile range (IQR) and the black whiskers reach the farthest point within 1.5 IQR of the end of the box.

(D) R^2 value across models.

identity, and eccentricity without confounding the variables. Additionally, likelihood ratio tests confirm that each model performs significantly better than the last, so one could use any model with confidence.

In many experiments, the sex and age of the animal is known. Therefore, our fifth model, sex and non-genetic (M5), adds sex and age as separate variables, reflecting the differences in body composition between sexes and as a mouse ages. This again provides a boost in accuracy and a reduction in error values. In certain conditions, the strain of the mouse is also known. Therefore, the sixth model, full (M6), introduces the strain of the mouse as a variable. Here, we include an interaction term between strain and sex because we observed that while males are generally heavier than females, there are some strains for which the opposite is true. The full model performs much better than the previous models, M1–M5, achieving a mean R^2 of 0.92 and a MAPE of 4.84% (Figure 3B). Visualizing the MAE and R^2 values of each model illustrates their relative improvements under cross-validation (Figures 3C and 3D). The performance values for each model are compared in Table 2.

Models and performance

Effect of sex and genetic diversity

We observe large differences in true and predicted mean body mass between sexes. In most, but not all, strains, males have a higher percentage of fat than females and are larger.[28] We tested the performance of the full model (M6) on males and females. While females weigh less than males on average, our model performs uniformly well across both females and males (Figure 4A). The differences in performance are slight: females and males have MAPEs of 4.82% and 5.17%, respectively.

Next, we tested the performance of the full model (M6) in genetically diverse mice and measured the accuracy and precision of the model across 44 classical inbred, 7 wild-derived inbred, and 11 F1 hybrid strains (Figure 4B). Most strains are between 20 and 30 g, with wild-derived strains like MSM/MsJ at the

Table 2. Model definitions and performance

Model	Model input variables	RMSE (g)	MAE (g)	MAPE (%)	R^2
Base (M1)	A_{px} (px^2)	2.729 ± 0.070	2.072 ± 0.052	8.584 ± 0.210	0.767 ± 0.012
M2	A_{cm} (cm^2)	2.469 ± 0.058	1.814 ± 0.042	7.528 ± 0.179	0.810 ± 0.010
M3	A_{cm} (cm^2), arena	2.416 ± 0.052	1.765 ± 0.035	7.314 ± 0.156	0.818 ± 0.009
Geometric (M4)	A_e (cm^2), arena	2.322 ± 0.052	1.739 ± 0.039	7.223 ± 0.174	0.832 ± 0.010
Sex and non-Genetic (M5)	A_e (cm^2), arena, sex, age	2.201 ± 0.065	1.622 ± 0.037	6.774 ± 0.167	0.849 ± 0.010
Full (M6)	A_e (cm^2), arena, sex, age, strain	1.606 ± 0.069	1.162 ± 0.031	4.836 ± 0.137	0.920 ± 0.007

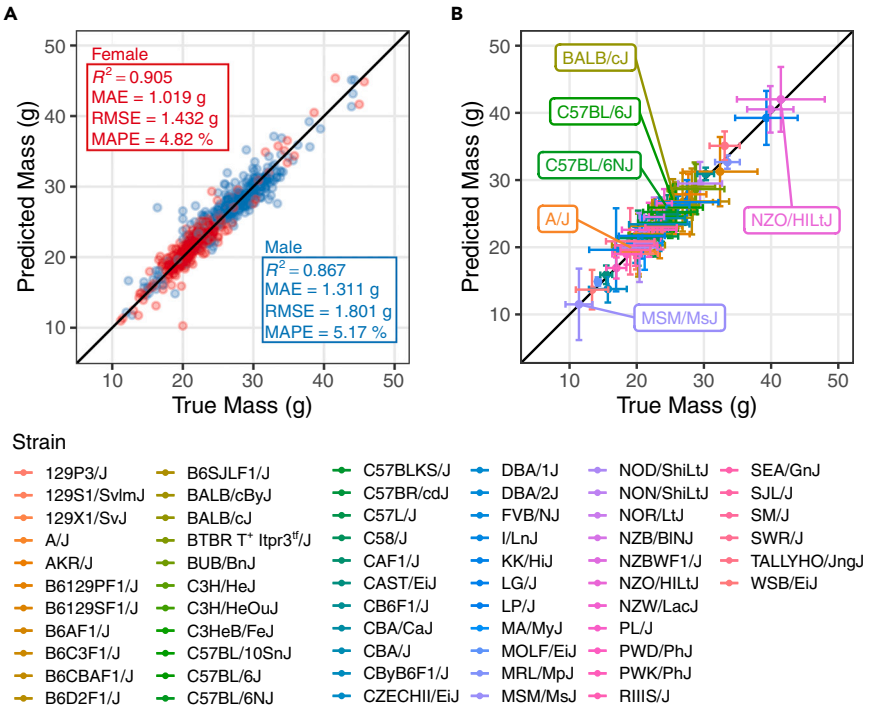


Figure 4. Full model (M6) performance by sex and strain

(A) Full model performance faceted by sex with R^2, MAE, RMSE, and MAPE values presented for males (in blue) and females (in red).

(B) Full model performance faceted by strain (mean ± standard deviation). The four strains from Figure 1B are labeled here, along with the strains with the smallest and the largest mean areas. The test and training data are the same as in Figure 3B for ease of comparison. The full results are presented in Table S1.

low end and obesity models such as NZO/HILtJ on the high end of the distribution (labeled in Figure 4B). We quantify and demonstrate the model's performance by averaging the predictions and calculating their variability across individual strains. The model accurately predicted the mean mass across highly genetically diverse mice (Figure 4B). Our model predictions also precisely capture the variability in observed weights across each strain (Figures 4B, error bars, S2, and S3). We quantify the mean observed and predicted mass and standard deviation for each strain in Table S1. Additionally, we retrain model M6 on our entire dataset and present the observed and predicted RSDs in mass for each strain, demonstrating that real strain-level variation is captured in our model (Figures S2 and S3). This final model can be used for inferring on any new data. We conclude that our model performs well across sex and diverse genetics.

Longitudinal tracking of mouse mass

We trained our visual mass prediction models using genetically diverse mouse strains and ages tested only once in the open field. This was a reasonable approach to build a model that generalizes to the visual diversity seen in the mouse. Next, we determined whether our approach could consistently track the weight of the same mouse over time. We reasoned that if a mouse suddenly lost weight or was under treatment that results in adverse effects, our non-invasive approach would have diagnostic utility. We manually assessed the body mass of 16 mice and then tested them in the 1-h open field over a 22-day period. Specifically, we tested 4 C57BL/6J females and 12 C57BL/6J males every day for 13 days, followed by an additional test day 9 days later (day 22). We then applied the previously trained M1–M6 models to these data and evaluated the performance using quantitative and qualitative methods. We found that the full model, M6, performed best in terms of MAE (1.54 ± 0.08), RMSE (3.51 ± 0.33), MAPE (5.4% ± 0.25%), and $\widehat{R}^2 = 0.87$ (Figures 5A and 5B). Qualitatively, we found that the visual prediction of the mass of each of the 16 animals closely mirrored the manually measured mass (Figure 5C). These results demonstrate that our models perform well when tracking the mass of individual animals over time.

Multi-day visual mass tracking with multiple mice in home environment

To extend the applicability of our methods, we further tested the performance of our models in a long-term housing environment with multiple mice. We collected video data in a modified open-field environment with bedding, food, and a water Lixit. This environment, the JAX Animal Behavior System (JABS), has been described in a preprint.[23] We housed three female C57BL/6J mice in this arena for 4 continuous days. The mass of each animal was measured manually at the beginning of the 4-day experiment and used as ground truth. We randomly selected 26 video clips of ~30 s or less for analysis from the active (dark) phase over a 4-day period. Since our visual mass models were trained on single mice, we selected 8 clips in which the mice were physically separated and distinct for analysis (pipeline summarized in Figure 6A). Each of these 8 clips is 10–20 s long and has a resolution of 800 × 800 px. The segmentation model used for single mice in the open field was trained on diverse mice in a plain open field without bedding, food, or water lixit.[20] It does not perform well in the home environment with increased complexity, and a new segmentation model was needed. To predict multi-mouse segmentation, we applied the track anything model (TAM) approach.[29] This is a video extension of the segment anything model (SAM)[30] that propagates segmentation predictions forward in time. We initialized the mice in a video using a manual keypoint prompt, which is typically 1 click per animal to indicate it as an object to track. After that, the model can predict the segmentations of that mouse for the remainder of the video. We visually inspected the performance of this model to ensure the quality of the segmentation. We confirmed that the identities of all mice remained consistent within a video by manually inspecting each frame of each video. We also manually matched the identities of individual mice between videos. The resulting segmentation mask for each mouse was used for modeling using M6.

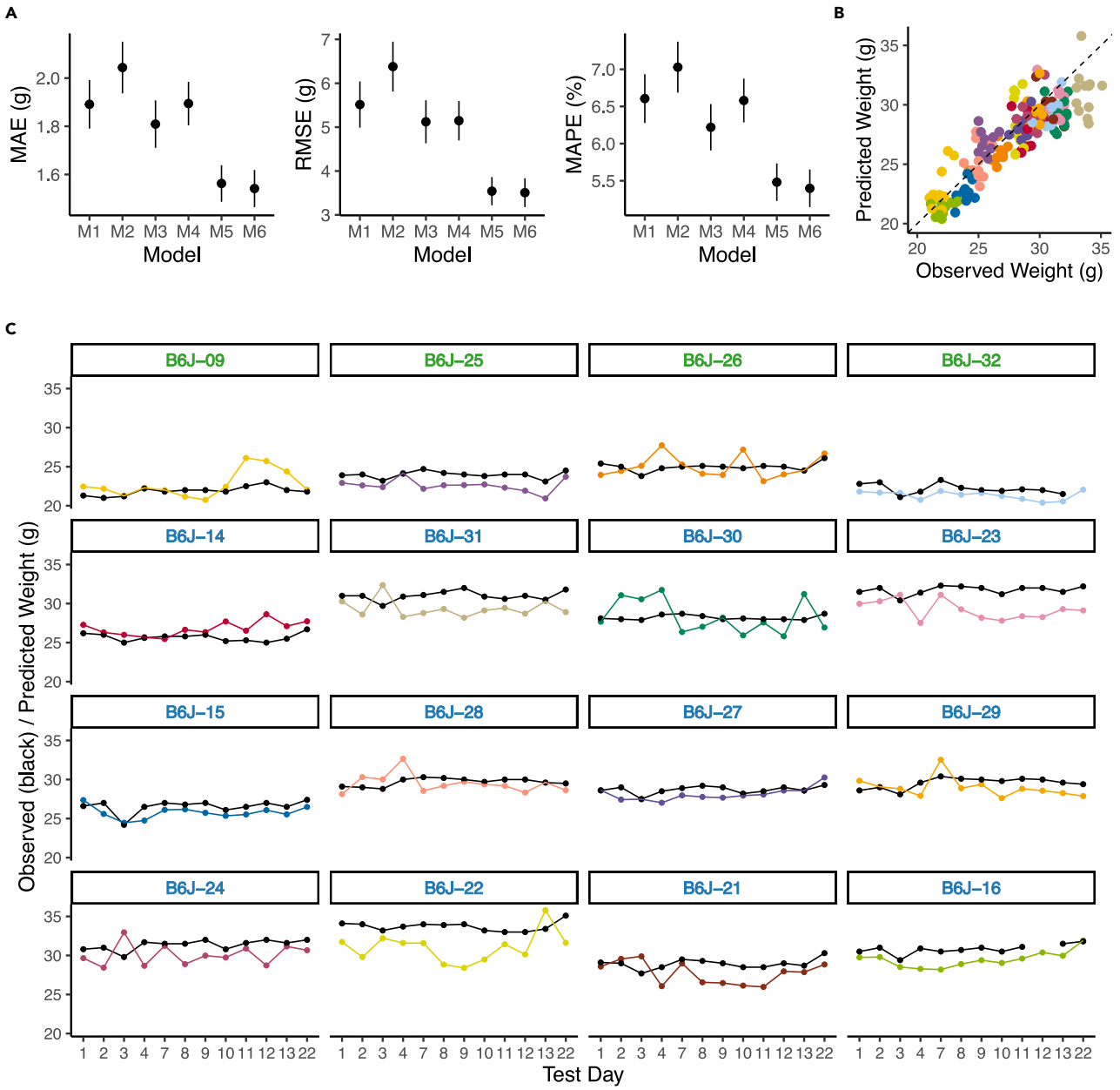


Figure 5. Successful longitudinal tracking of individual mouse mass for 22 days in the open field

Mouse masses were manually (black) and visually (color) assessed over a 22-day period in the 1-h open field for the same animal.

(A) Comparison of performance (mean ± standard error) of previously trained models on prediction of mass on new data from a longitudinal experiment. We found that M6 had the lowest error (MAE, RMSE, and MAPE).

(B) A scatterplot comparison of predicted and observed weights of the 16 mice (matching colors in B and C, $\widehat{R}^2 = 0.87$). Points on the dashed diagonal line indicate perfect predictions.

(C) Individual mouse data on each day. Plots of predicted weights (color) to observed weights (black) across each mouse (panels) for each day of the experiment. Graph title panel text color indicates sex of the animal (females are green and males are blue).

We find that M6 has very good performance in this new environment and, on average, closely follows the true mass over time in each video (Figure 6B, solid colored lines represent the median mass prediction in the time interval). It achieves a median error below 10% in all three mice: mouse 1 achieves a 3.15% mean error, mouse 2 achieves a 6.41% mean error, and mouse 3 achieves a 9.73% mean error (Figure 6C). Upon inspection of the videos, we observed that the slightly higher error in mouse 3 appears to be due to its behavior. This mouse frequently jumps onto the walls and onto the lixit. Regardless, these error levels

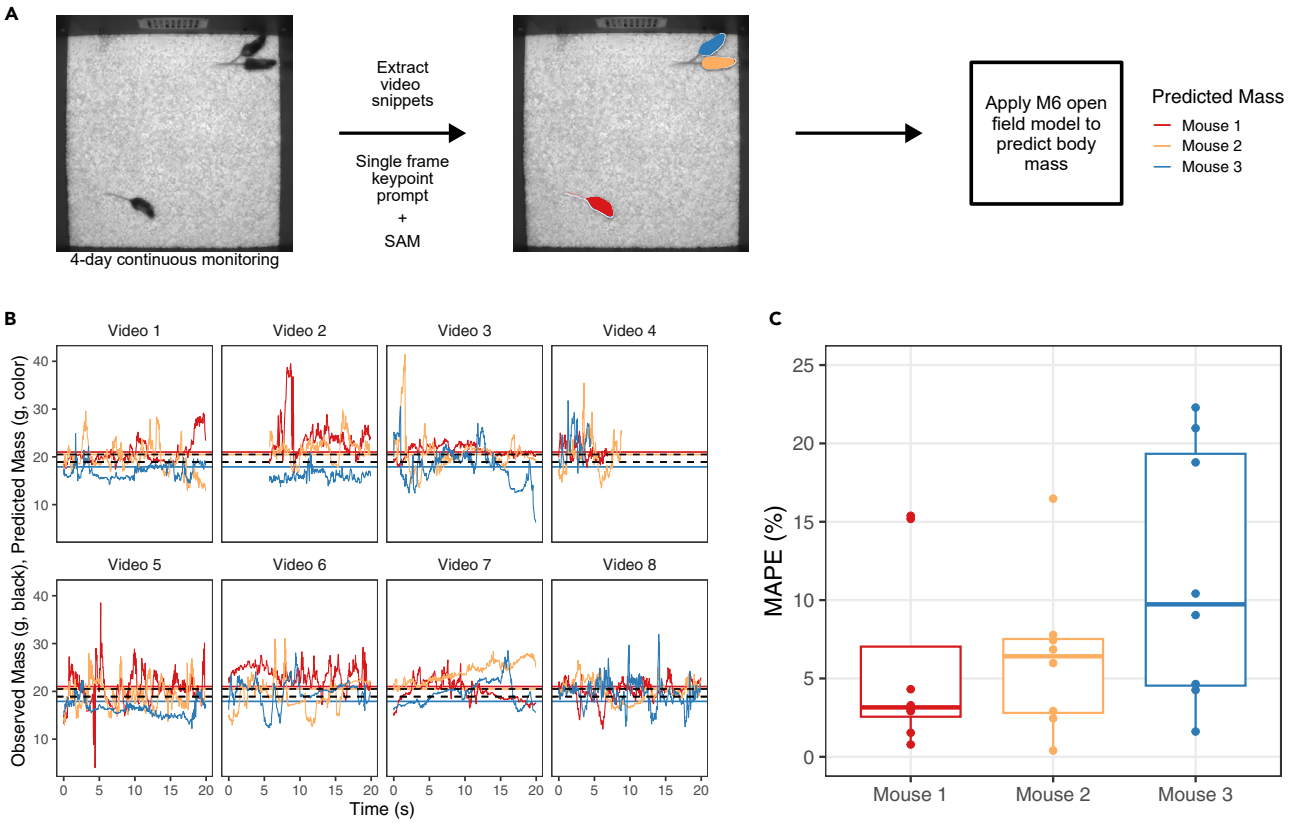


Figure 6. Successful visual mouse mass assessment over a multi-day experiment with multiple mice in a home environment

(A) Three mice were cohoused with bedding, food, and water for 4 continuous days. We used the TAM to segment each mouse and applied M6 to visually predict mouse mass.

(B) We selected 8 video clips and predicted mass visually in every frame using M6 (colored lines). Each mouse's median predicted mass is shown as a colored horizontal line, and the true masses are shown as two black dashed lines (two mice had equal measured masses). The manual mass was measured at the start of the experiment.

(C) MAPE of mass prediction for each of the three mice over the 8 videos. Each box shows the interquartile range (IQR) and the black whiskers reach the farthest point within 1.5 IQR of the end of the box.

are comparable to the observed performance in the 1-h open field, and we conclude that our method can be applied toward long-term home monitoring conditions.

DISCUSSION

Body mass is an important measure of health in mice and is widely used as a feature of both veterinary intervention and behavioral and drug trial experiments. However, frequently monitoring body mass by hand is impractical over a long time period and, more importantly, can negatively affect the actual experiment being performed. We sought to test whether visual methods provide adequate accuracy and precision to be a viable alternative to manual weighing animals. Mice are highly deformable and change their posture in very short timescales. In addition, since our camera is positioned relatively high above the mouse, only 0.6% of the pixels in the video frame belong to it. Thus, we do not have a very high-resolution view of a highly deformable object. To solve this problem, we applied computer vision methods to develop models that predict the body mass of an individual mouse in an open-field arena. We developed six models for this, each using a selection of visual information extracted from the segmentation and ellipse-fit masks, the particular arena a mouse was recorded in, and biological attributes of the mouse such as age, sex, and strain.

There are several advantages to our approach. First, our approach uses an explainable machine learning approach. The first step of our approach is to generate segmentation masks for the mouse. Segmentation is highly flexible and one of the oldest and most established computer vision tasks, particularly for biomedical images.[31,32] We selected a segmentation step to enable observable detection for accuracy shifts. For example, if a mouse escapes the arena, we can detect that no segmentation is predicted and make no body mass prediction.

Modern methods of segmentation improve upon traditional methods by using neural networks.[29,30,33–36] Segmentation neural networks can be trained using much lower training data, and there are already pretrained models for general purpose segmentation. For instance, we trained a segmentation network for sleep-state prediction using 313 frames.[37] The creation of training data in different visual environments should be much easier with new, more general segmentation foundation models

such as the SAM[30] and TAM[29] and self-supervised models such as DINOv2.[38] These models seem to work well for segmentation of animals with minimal supervision as demonstrated here. Indeed, in our home environment experiment, we were able to segment mice with minimal supervision using the TAM. After segmentation, we apply well-understood and explainable techniques including unit transformations, normalization, and linear regression modeling. Thus, our method is flexible for adoption to new environments and organisms without the burden of creating large training data while producing explainable results. We envision that such a system can be used to assess mass in a changing table that is routinely used in mouse facilities.

A second advantage of our approach is the use of highly diverse mouse strains, with varied coat colors, sizes, and behaviors, for training our models. This ensures that the model can handle diverse visual and size distributions that may be seen in real-world settings. Each of our six models has different experimental applications. Though it is clear that the full model has the highest accuracy and lowest error, it is not necessarily suitable for all experiments. Practically, one would apply the geometric, non-genetic, or full model to an experiment, not M1, M2, or M3. While in some sense, M1 is the most general of all the models, there is no information in the geometric model (M4) that is not easily attainable. Its two variables, A_e and arena, are derived from the same base information in the video, not from any further information about the mouse itself. For the same reasons, it would equally make sense to use M4 instead of M2 or M3. As mentioned in the previous section, it is still useful to present M1–M3 because they show the benefit of additions and modifications to the models.

With fine-tuning to the particular environment, the geometric model (M4) can be applied to all open-field videos of similar size and resolution with an individual mouse. However, it would be a rare mouse model experiment that intentionally did not take into account the sex and age of the mice, as these are often important variables. The sex and non-genetic model (M5), as its name suggests, would be very useful for experiments where collecting strain information is unfeasible or undesirable. A key example of this is experiments that use genetically diverse populations of mice, such as diversity outbred or other genetically heterogeneous populations that are often used for quantitative genetic studies.[39,40] The full model (M6), which includes genetic information, is readily applicable in studies in a defined genetic background. Our full model would provide the highest accuracy monitoring of body mass in these areas. Furthermore, any classical mouse experiment with inbred strains that lends itself to having one mouse in an open-field environment would benefit from this model, especially behavioral experiments where inducing anxiety potentially introduces bias.

To avoid drawing erroneous conclusions about medically concerning fluctuations in body mass, this type of visual mass prediction is helpful for health monitoring only if its error is well within the bounds of human intervention points. Following the IACUC Humane Intervention guidelines, it is necessary to intervene, typically with euthanasia, if a mouse loses more than 20% of its body mass compared to similar control animals. As an example, if an inbred mouse lost 20% of its body mass during an experiment, then our results indicate that we could use M6, the full model, to predict that the mouse lost ∼15.2%–24.8% of its body mass, which is a small enough range that we know something is likely wrong with it and should be investigated further. In this way, our visual approach can serve well as a wide-screen diagnostic tool for health monitoring in its current state. For studies that do not require adverse event detection, the sensitivity of our method has added value. For instance, a drug or genetic manipulation that leads to slight (∼5%) change in mass can be detected accurately, as demonstrated by our multi-day individual tracking. Although a 5% change in mass may seem low, in human clinical trials for weight loss drugs, the number of participants who achieve at least a 5% decrease in mass is the primary result that is reported.[41] This level of sensitivity makes this a useful tool beyond health checks for preclinical animal studies.

Here, our goal is to quantify the accuracy and precision of visual mass determination using an existing video acquisition system without optimizing the data stream for mass determination. Most labs collect video for behavior analysis, and we reason that if a video stream has enough information to quantify behavior, then it should be adequate for mass determination. We demonstrate that the error using the existing video is within tolerance to meet the IACUC requirements of mass change and provide a viable alternative to manual mass assessment. While our models achieve good performance, they are not without error. These errors can be broken down into different types based on their source of introduction: equipment spatial resolution, limitations of our selected measurement, and observation interval. Since we image at a spatial resolution of 480 × 480, where the mouse occupies a small percentage (<1% of area) of the visible area, increasing the resolution of the image should reduce the error. Our measurement is that of the silhouette of the mouse from a top-down viewing angle. This alone does not capture the complex posture of the animal. Using calibrated depth cameras to provide a better understanding of the animal's volume would reduce the prediction error. Such cameras have been used in industrial farming[17,18] and unsupervised behavior analysis of rodents.[42] Finally, the observation interval limits how much information can be used to make a prediction. Ideally, predictions could be made with a couple of frames. However, we observed that there is high variation in area across the course of an open-field assay, as the mouse performs different behaviors. This variation necessitates longer observational intervals to ensure accurate body mass predictions. Yet, in our home cage experiment, we achieved a low predictive error using only ∼20-s intervals, which suggests that functional predictions can be made on shorter time frames.

Future development of this method can expand into continuous monitoring in shoebox-size home cages. Home cage settings can vary widely between experimenters and are not standardized, making visual prediction of body mass a greater challenge.[43–45] However, many experiments are better suited to social housing, and visually tracking body mass in home cages would be very useful in production colonies. Our home environment results using the SAM show that with the addition of more general segmentation techniques, our models can be applied to multiple mice housed together in more complex environments. Furthermore, the camera may be much closer to the animal in home cage environments and may also contain perspective, distortion, occlusion, and even behavioral challenges such as huddling and nesting. Our methods can be extended to handle these conditions in

the future. Our methods can also be extended to other organisms such as rats and non-human primates, where manual mass assessment is much harder.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Vivek Kumar (vivek.kumar@jax.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Original video data have been deposited to Harvard Dataverse: https://doi.org/10.7910/DVN/SAPNJG with the "mass" tag.[46] JAX is the owner of this dataset and has openly released it under the "CC BY-NC-SA" license. This license is described in detail here (https://creativecommons.org/licenses/by-nc-sa/4.0/) and allows non-commercial use and reuse (share alike) of the data with proper attribution.

The segmentation neural network code and weights have been previously published and are available.[20] All code is publicly available through the Kumar Lab GitHub (https://github.com/KumarLabJax/visual-mouse-weight/tree/v1.1.0). The version of code used in this paper is archived on Zenodo.[47]

Mice

All mice were obtained from The Jackson Laboratory production colonies. The mouse videos were obtained from a previously conducted strain survey of over 2,400 mice from 62 strains, including at least 4 male and 4 for almost every strain[20,48] (see Table S1). We selected 2,028 individual mice out of that survey, leaving out those with missing features in the data like age or arena recorded. Our mice vary in age from 7 to 81 weeks old, with a mean age of 11 weeks. Each mouse was weighed with an analytical laboratory scale (Ohaus) immediately prior to recording. The sampled body weights range from 9.1 to 54 g with a mean of 24.7 g.

Arena recording setup

All behavior methods and datasets have been previously described.[20,48,49] We have published detailed descriptions of our data acquisition methods, hardware, and software.[23] Briefly, all videos were recorded at 30 fps, have 8-bit monochrome depth, run for 55 min, and are saved as 480 × 480 px files. Each camera was mounted approximately 100 cm above each arena, with zoom settings tuned to 8 px/cm. Variations in this zoom were normalized by the corner detection approach described in the results.[27] We recorded individual mice in six near-identical open-field arenas. The arenas are 52 × 52 × 23 cm and built with white PVC plastic floors and gray PVC plastic walls. A white 2.54-cm chamfer was added to all inner edges for easier cleaning. Each arena was illuminated by a light-emitting diode light producing 200–600 lux of light.

Statistical analysis

To predict the weight of the mouse, we fit two simple linear regression models (M1, M2) and four multiple linear regression models (M3, M4, M5, M6) using video and animal-specific covariates. We describe our model as follows:

$$y_i = \beta_0 + \beta_1 \text{area}_i + \beta_2 \text{age}_i + \beta_3 \text{sex}_i + \sum_{b=1}^{B} \beta_b^{\text{arena}} \text{arena}_{ib} + \sum_{s=1}^{S} \beta_s^{\text{strain}} \text{strain}_{is} + \sum_{s=1}^{s} \gamma_{is} \text{sex}_i \times \text{strain}_{is} + \epsilon_i, \epsilon_i \sim \mathcal{N}(0, \sigma^2),$$

where y_i denotes the weight of the mouse i; $\beta_p, p = 1,2,3$, β_b^{arena}, and β_s^{strain} are the regression coefficients associated with area_i, age_i, sex_i, arena_i, and strain_i covariates for mouse i; γ_i represents the interaction effect between strain and sex; and ϵ_i denotes the error term. The regression coefficients β and γ are estimated from the data using the least-squares algorithm from the data so as to use the model to predict the weight of a new mouse. Further,

$$\text{strain}_{is} = \begin{cases} 1 & \text{if mouse } i \text{ belongs to strain s} \\ 0 & \text{if mouse } i \text{ does not belong to strain s} \end{cases}$$

and

$$\text{arena}_{ib} = \begin{cases} 1 & \text{if mouse } i \text{ belongs to arena b} \\ 0 & \text{if mouse } i \text{ does not belong to arena b} \end{cases}$$

Each of our six models (M1–M6) can also be conveniently described as follows:

$$\text{M1} : \text{weight} \sim A_{px},$$

$$\text{M2} : \text{weight} \sim A_{cm},$$

$$\text{M3} : \text{weight} \sim A_{cm} + \text{arena},$$

$$\text{M4} : \text{weight} \sim A_e + \text{arena},$$

$$\text{M5} : \text{weight} \sim A_e + \text{arena} + \text{age} + \text{sex}, \text{and}$$

$$\text{M6} : \text{weight} \sim A_e + \text{arena} + \text{age} + \text{sex} + \text{strain} + \text{sex} * \text{strain}.$$

To assess the accuracy of our model predictions, we split the data randomly into two parts: train (70%) and test (30%). The test set served as an independent evaluation sample for the models' predictive performance. We performed 50-fold cross-validation to allow for a proper assessment of uncertainty in our test set results. The models were compared in terms of MAE, MAPE, RMSE, and R^2. These are defined as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|,$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2},$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%, \text{and}$$

$$R^2 = \text{Cor}(y, \hat{y})^2,$$

where y and $\hat{y}$ are the true (observed) and predicted values from the model, respectively.

We also performed likelihood ratio tests between each successive model that confirmed that each higher-complexity model provides a higher goodness of fit than the previous model, i.e., M2 provides a significantly higher goodness of fit than M1, and so on. We fitted our models in R and performed statistical tests with the lmtest and AICcmodavg packages.

Segmentation and modeling in the home environment

To predict multi-mouse segmentation, we apply the TAM approach.[29] This is a video extension of SAM[30] that propagates segmentation predictions forward in time. We initialize the mice in a video using a manual keypoint prompt, which is typically 1 click per mouse to indicate it as an object to track. After that, the model can predict segmentations for that mouse for the remainder of the video. We discard frame-wise prediction data when the predicted segmentation masks no longer predict on the mouse. Additionally, we discard entire videos when mice were huddling or occluded because obtaining accurate segmentation in these difficult situations is beyond the scope of this paper.

Our segmentation method used for single mouse experiments, which our mass prediction models are trained on, includes a 1-px border around the mouse. The SAM does not predict this border. Additionally, this long-term data are imaged at 1.7× the resolution of the open-field experiments (see experimental procedures). As such, we add a 2-px border to the raw TAM predictions by applying two 3 × 3 dilation morphological filters. We chose to use 20-s videos (as opposed to the 55-min videos used in our open-field modeling) because the TAM is RAM hungry. Larger videos would cause out-of-memory errors, and

tracking tends to fall apart before then. We extracted the same visual parameters previously described from this dilated segmentation mask and fed them into M6, just as before. We only used one cage in this experiment, so unit conversion can be done manually without using the corner detection needed in the open-field experiments.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2024.101039.

ACKNOWLEDGMENTS

We thank Kumar Lab members for helpful advice and comments. We thank Kayla Dixon, a Colby College Lunder Fellow in the Kumar Lab, who initiated the project. We thank Marina Santos, Tom Sproule, and Sean Deats for behavioral testing. This work was funded by The Jackson Laboratory Directors Innovation Fund, National Institutes of Health, National Institute of Drug Abuse DA051235 and DA048634 (NIDA; V.K.), National Institute on Aging AG078530 (NIA; V.K.), National Institute of Neurological Disorders and Stroke NS078795 (NINDS; M.G.), and the C.C. Little Scholarship Fund (M.G.).

AUTHOR CONTRIBUTIONS

M.G., B.G., and V.K. designed the experiments and analyzed the data. G.S. and M.G. carried out statistical modeling analysis. All authors wrote and edited the paper.

DECLARATION OF INTERESTS

The Jackson Laboratory has filed a patent on the methods described here.

Received: August 10, 2023
Revised: February 20, 2024
Accepted: July 11, 2024
Published: August 7, 2024

REFERENCES

1. Eknoyan, G. (2008). Adolphe quetelet (1796–1874)—the average man and indices of obesity. Nephrol. Dial. Transplant. *23*, 47–51.
2. Negri, E., Pagano, R., Decarli, A., and La Vecchia, C. (1988). Body weight and the prevalence of chronic diseases. J. Epidemiol. Community Health *42*, 24–29.
3. Knight, J.A. (2011). Diseases and disorders associated with excess body weight. Ann. Clin. Lab. Sci. *41*, 107–121.
4. Vandamme, T.F. (2014). Use of rodents as models of human diseases. J. Pharm. BioAllied Sci. *6*, 2–9.
5. Ahloy-Dallaire, J., Klein, J.D., Davis, J.K., and Garner, J.P. (2019). Automated monitoring of mouse feeding and body weight for continuous health assessment. Lab. Anim. *53*, 342–351.
6. Hankenson, F.C. (2014). Critical Care Management for Laboratory Mice and Rats (Taylor & Francis Group), pp. 25–42.
7. Talbot, S.R., Biernot, S., Bleich, A., van Dijk, R.M., Ernst, L., Hager, C., Helgers, S.O.A., Koegel, B., Koska, I., Kuhla, A., et al. (2020). Defining body-weight reduction as a humane endpoint: a critical appraisal. Lab. Anim. *54*, 99–110.
8. Hurst, J.L., and West, R.S. (2010). Taming anxiety in laboratory mice. Nat. Methods *7*, 825–826.
9. Meijer, M., Sommer, R., Spruijt, B., van Zutphen, L., and Baumans, V. (2007). Influence of environmental enrichment and handling on the acute stress response in individually housed mice. Lab. Anim. *41*, 161–173.
10. Dohmen, R., Catal, C., and Liu, Q. (2021). Computer vision-based weight estimation of livestock: a systematic literature review. New Zealand Journal of Agricultural Research *65*, 227–247.
11. Fernandes, A.F.A., Dórea, J.R.R., and de Magalhães Rosa, G.J. (2020). Image analysis and computer vision applications in animal sciences: An overview. Front. Vet. Sci. *7*, 551269.
12. Nyalala, I., Okinda, C., Kunjie, C., Korohou, T., Nyalala, L., and Chao, Q. (2021). Weight and volume estimation of poultry and products based on computer vision systems: a review. Poult. Sci. *100*, 101072.
13. Hui, L., Qiang, X., Shilin, L., Libin, Z., and Hongsheng, Y. (2015). Evaluation of body weight of sea cucumber apostichopus japonicus by computer vision. Chin. J. Oceanol. Limnol. *33*, 114–120.
14. Ruby, J.G., Di Francesco, A., Ylagan, P., Luo, A., Keyser, R., Williams, O., Spock, S., Li, W., Vongtharangsy, N., Chatterjee, S., et al. (2023). An automated, home-cage, video monitoring-based mouse frailty index detects age-associated morbidity in c57bl/6 and diversity outbred mice. J. Gerontol. A Biol. Sci. Med. Sci. *78*, 762–770.
15. Kamchen, S.G., dos Santos, E.F., Lopes, L.B., Vendrusculo, L.G., and Condotta, I.C. (2021). Application of depth sensor to estimate body mass and morphometric assessment in nellore heifers. Livest. Sci. *245*, 1–10.
16. Okayama, T., Kubota, Y., Toyoda, A., Kohari, D., and Noguchi, G. (2021). Estimating body weight of pigs from posture analysis using a depth camera. Anim. Sci. J. *92*, e13626.
17. Arulmozhi, E., Bhujel, A., Moon, B.-E., and Kim, H.-T. (2021). The application of cameras in precision pig farming: An overview for swine-keeping professionals. Animals. *11*, 2343.
18. Oliveira, D.A.B., Pereira, L.G.R., Bresolin, T., Ferreira, R.E.P., and Dorea, J.R.R. (2021). A review of deep learning algorithms for computer vision systems in livestock. Livest. Sci. *253*, 104700.
19. Hakem, M., Boulouard, Z., and Kissi, M. (2022). Classification of body weight in beef cattle via machine learning methods: A review. Procedia Computer Science *198*, 263–268.
20. Geuther, B.Q., Deats, S.P., Fox, K.J., Murray, S.A., Braun, R.E., White, J.K., Chesler, E.J., Lutz, C.M., and Kumar, V. (2019). Robust mouse tracking in complex environments using neural networks. Commun. Biol. *2*, 124.
21. van Gaalen, M.M., and Steckler, T. (2000). Behavioural analysis of four mouse strains in an anxiety test battery. Behav. Brain Res. *115*, 95–106.
22. Laarakker, M.C., van Lith, H.A., and Ohl, F. (2011). Behavioral characterization of a/j and c57bl/6j mice using a multidimensional test: association between blood plasma and brain magnesium-ion concentration with anxiety. Physiol. Behav. *102*, 205–219.
23. Beane, G., Geuther, B.Q., Sproule, T.J., Trapszo, J., Hession, L., Kohar, V., and Kumar, V. (2022). Video based phenotyping platform for the laboratory mouse. Preprint at bioRxiv. https://doi.org/10.1101/2022.01.13.476229.
24. Shcherbakov, M.V., Brebels, A., Shcherbakova, N.L., Tyukov, A.P., Janovsky, T.A., and Kamaev, V.A. (2013). A survey of forecast error measures. World Appl. Sci. J. *24*, 171–176.
25. De Myttenaere, A., Golden, B., Le Grand, B., and Rossi, F. (2016). Mean absolute percentage error for regression models. Neurocomputing *192*, 38–48.
26. Armstrong, J.S., and Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. Int. J. Forecast. *8*, 69–80.
27. Sheppard, K., Gardin, J., Sabnis, G.S., Peer, A., Darrell, M., Deats, S., Geuther, B., Lutz, C.M., and Kumar, V. (2022). Stride-level analysis of mouse open field behavior using deep-learning-based pose estimation. Cell Rep. *38*, 110231.
28. Reed, D.R., Bachmanov, A.A., and Tordoff, M.G. (2007). Forty mouse strain survey of body composition. Physiol. Behav. *91*, 593–600.
29. Yang, J., Gao, M., Li, Z., Gao, S., Wang, F., and Zheng, F. (2023). Track anything: Segment anything meets videos. Preprint at arXiv. https://doi.org/10.48550/arXiv.2304.11968.
30. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., et al. (2023). Segment anything. Preprint at arXiv. https://doi.org/10.48550/arXiv.2304.02643.

31. Pal, N.R., and Pal, S.K. (1993). A review on image segmentation techniques. Pattern Recogn. *26*, 1277–1294.

32. Pham, D.L., Xu, C., and Prince, J.L. (2000). Current methods in medical image segmentation. Annu. Rev. Biomed. Eng. *2*, 315–337.

33. Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., and Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. IEEE Trans. Pattern Anal. Mach. Intell. *44*, 3523–3542.

34. Zou, X., Yang, J., Zhang, H., Li, F., Li, L., Gao, J., and Lee, Y.J. (2023). Segment everything everywhere all at once. Preprint at arXiv. https://doi.org/10.48550/arXiv.2304.06718.

35. Butoi, V.I., Ortiz, J.J.G., Ma, T., Sabuncu, M.R., Guttag, J., and Dalca, A.V. (2023). Universeg: Universal medical image segmentation. Preprint at arXiv. https://doi.org/10.48550/arXiv.2304.06131.

36. Ma, J., and Wang, B. (2023). Towards foundation models of biological image segmentation. Nat. Methods *20*, 953–955.

37. Geuther, B., Chen, M., Galante, R.J., Han, O., Lian, J., George, J., Pack, A.I., and Kumar, V. (2022). High-throughput visual assessment of sleep stages in mice using machine learning. Sleep *45*, zsab260.

38. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al. (2023). Dinov2: Learning robust visual features without supervision. Preprint at arXiv. https://doi.org/10.48550/arXiv.2304.07193.

39. Svenson, K.L., Gatti, D.M., Valdar, W., Welsh, C.E., Cheng, R., Chesler, E.J., Palmer, A.A., McMillan, L., and Churchill, G.A. (2012). High-resolution genetic mapping using the mouse diversity outbred population. Genetics *190*, 437–447.

40. Threadgill, D.W., Miller, D.R., Churchill, G.A., and de Villena, F.P.-M. (2011). The collaborative cross: a recombinant inbred mouse population for the systems genetic era. ILAR J. *52*, 24–31.

41. Khera, R., Murad, M.H., Chandar, A.K., Dulai, P.S., Wang, Z., Prokop, L.J., Loomba, R., Camilleri, M., and Singh, S. (2016). Association of pharmacological treatments for obesity with weight loss and adverse events: a systematic review and meta-analysis. JAMA *315*, 2424–2434.

42. Wiltschko, A.B., Johnson, M.J., Iurilli, G., Peterson, R.E., Katon, J.M., Pashkovski, S.L., Abraira, V.E., Adams, R.P., and Datta, S.R. (2015). Mapping sub-second structure in mouse behavior. Neuron *88*, 1121–1135.

43. Voikar, V., and Gaburro, S. (2020). Three pillars of automated home-cage phenotyping of mice: novel findings, refinement, and reproducibility based on literature and experience. Front. Behav. Neurosci. *14*, 575434.

44. 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.

45. Richardson, C.A. (2015). The power of automated behavioural homecage technologies in characterizing disease progression in laboratory mice: A review. Appl. Anim. Behav. Sci. *163*, 19–27.

46. Geuther, B., Deats, S., and Kumar, V. (2024). Jax - kumar lab mouse strain survey open field video dataset. Harvard Dataverse. https://doi.org/10.7910/DVN/SAPNJG.

47. Guzman, M., Geuther, B., Sabnis, G., and Kumar, V. (2024). Code for this article, "Highly Accurate and Precise Determination of Mouse Mass Using Computer Vision". kumarlabjax/visual-mouse-weight: v1.1.0. Zenodo. https://doi.org/10.5281/zenodo.11106337.

48. Geuther, B.Q., Peer, A., He, H., Sabnis, G., Philip, V.M., and Kumar, V. (2021). Action detection using a neural network elucidates the genetics of mouse grooming behavior. Elife *10*, e63207.

49. Kumar, V., Kim, K., Joseph, C., Thomas, L.C., Hong, H., and Takahashi, J.S. (2011). Second-generation high-throughput forward genetic screen in mice to isolate subtle behavioral mutants. Proc. Natl. Acad. Sci. USA *108*, 15557–15564.