Cell Reports Methods, Volume *3*

Supplemental information

LabGym: Quantification of user-defined animal behaviors using learning-based holistic assessment

Yujia Hu, Carrie R. Ferrario, Alexander D. Maitland, Rita B. Ionides, Anjesh Ghimire, Brendon Watson, Kenichi Iwasaki, Hope White, Yitao Xi, Jie Zhou, and Bing Ye

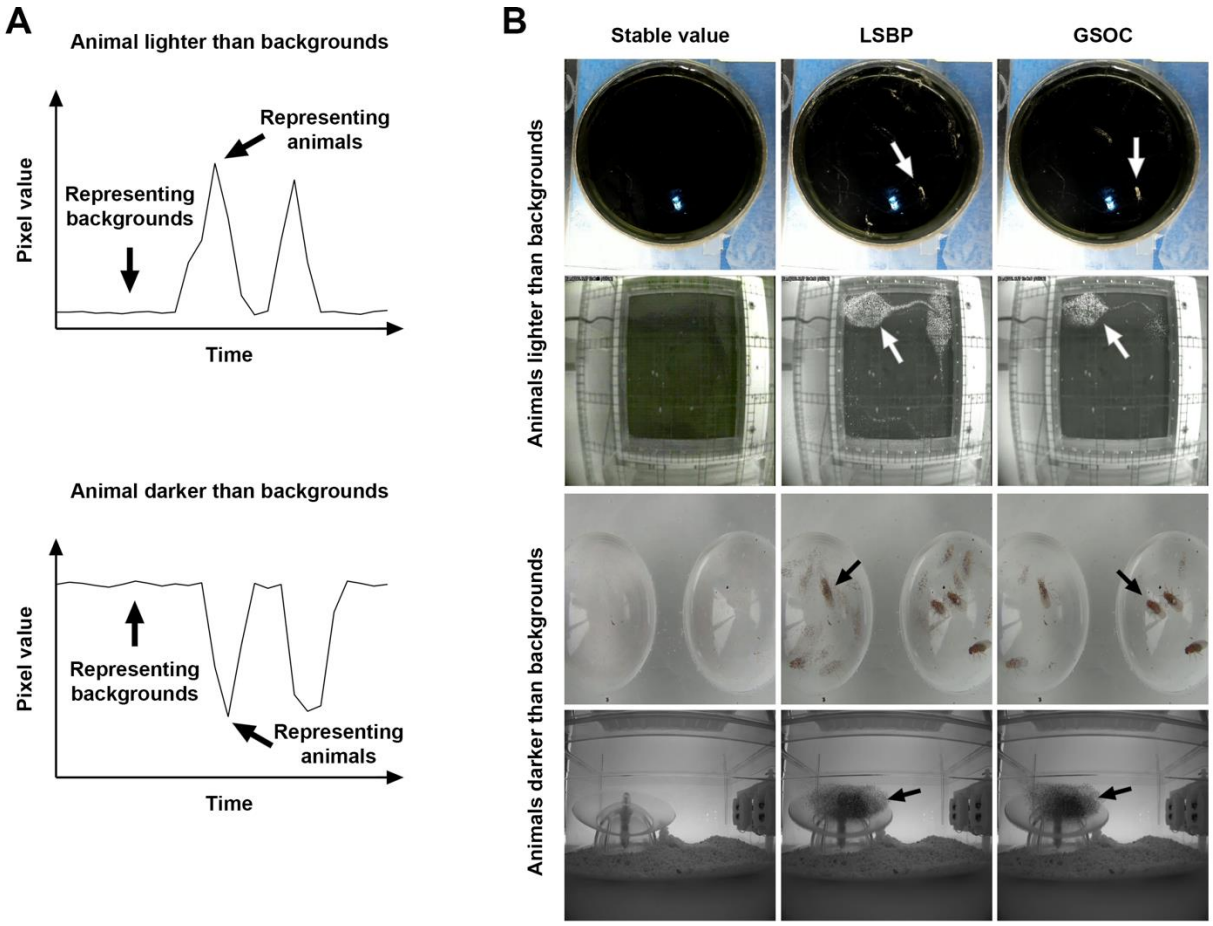

Figure S1. The stable-value detection method outperforms the state of the art, related to Figure 2. **A**. Illustrations showing the designing rationale of stable-value detection method for reconstructing the static background of a video in two different scenarios (animal lighter or darker than the backgrounds). **B**. Examples for reconstructed static backgrounds of the same videos using different methods. Arrows point at the remaining animal traces in the reconstructed static backgrounds.

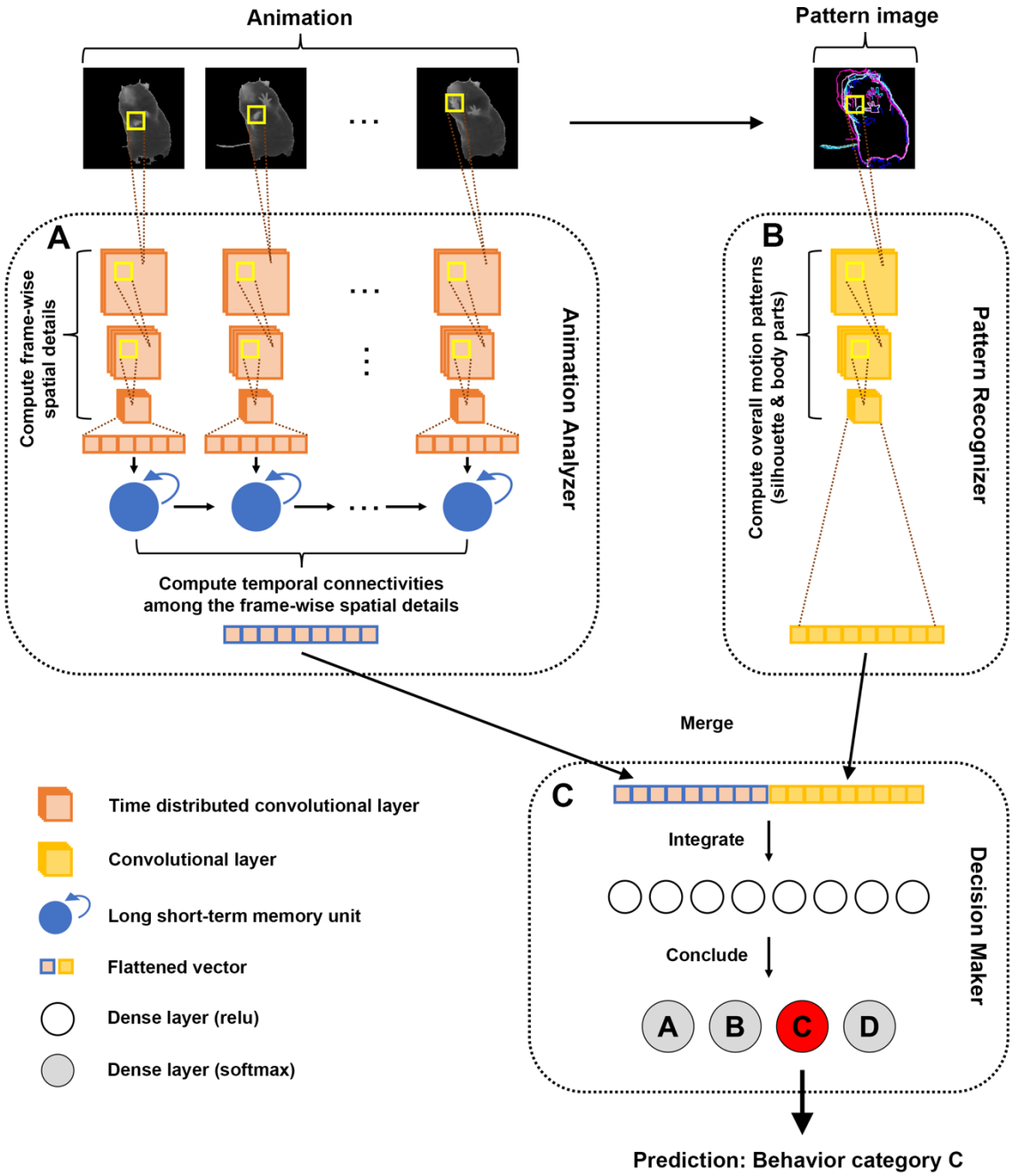

Figure S2. The design of the Categorizer, related to Figure 3.
A. The Animation Analyzer first uses time-distributed convolutional layers to compute all frame-wise spatial details of an animation, and then uses recurrent layers (long short-term memory, LSTM) to compute the temporal connectivity among the frame-wise spatial details.
B. The Pattern Recognizer uses convolutional layers to analyze the pattern image, which is superimposed contours of animal body (and body parts) along the temporal sequence (indicated by gradually changing colors) during a behavior.
C. The Decision Maker uses a concatenating layer to integrate the outputs from both the Animation Analyzer and the Pattern Recognizer, and then passes the integrated information to dense layers for concluding the behavioral categories.

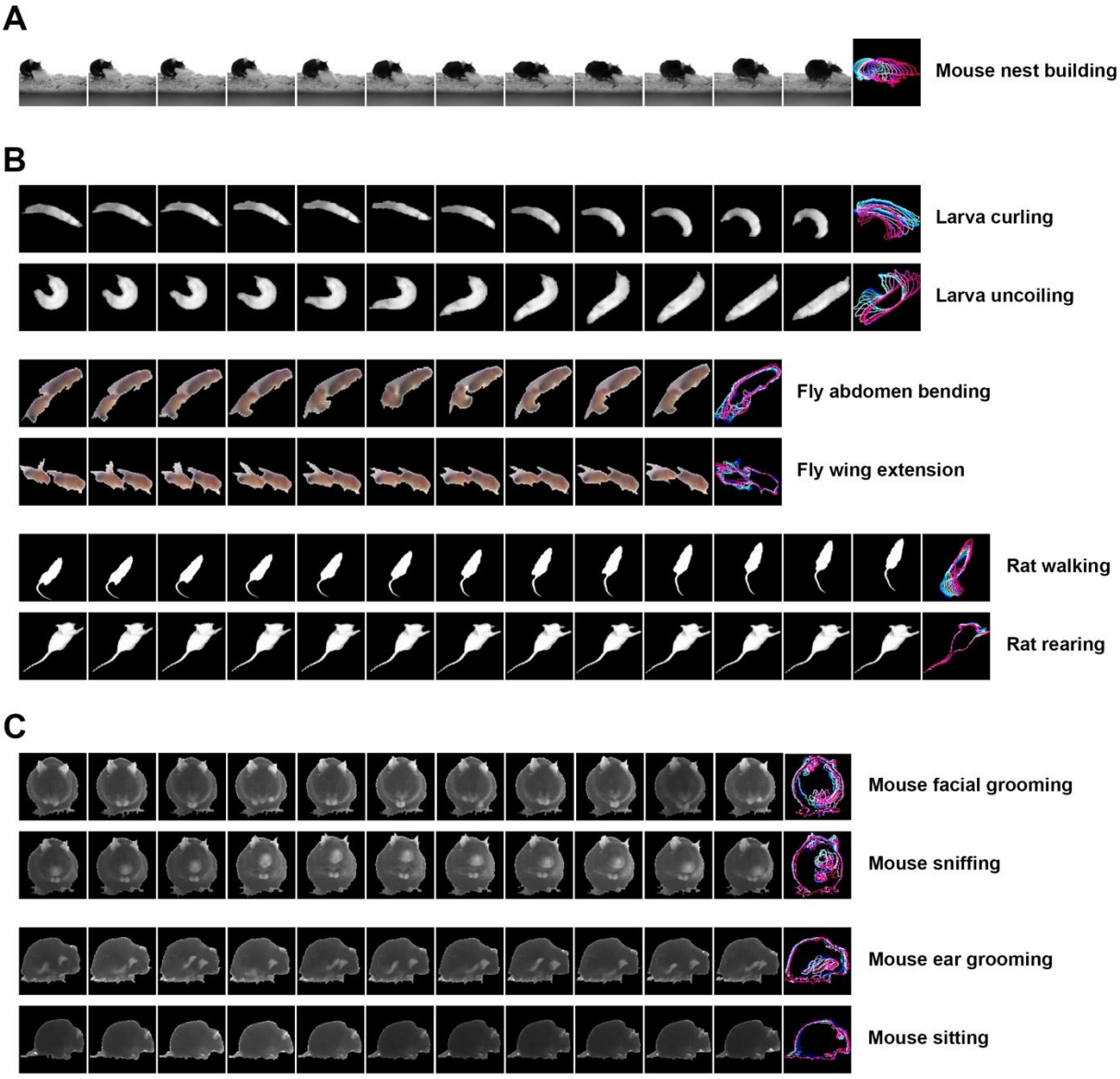

Figure S3. *LabGym* generates stand-alone, visualizable behavior examples, related to Figure 4.
A. Every frame of an animation (background is included) and its paired pattern image, showing nest building behavior of a mouse.
B. Every frame of animations and their paired pattern images (body parts are not included), showing (from top to bottom) larva curling and uncoiling, fly abdomen bending (copulation attempt) and wing extension (courtship song), and rat walking and rearing.
C. Every frame of animations and their paired pattern images (body parts are included), showing (from top to bottom) mouse facial grooming, sniffing, ear grooming, and sitting.

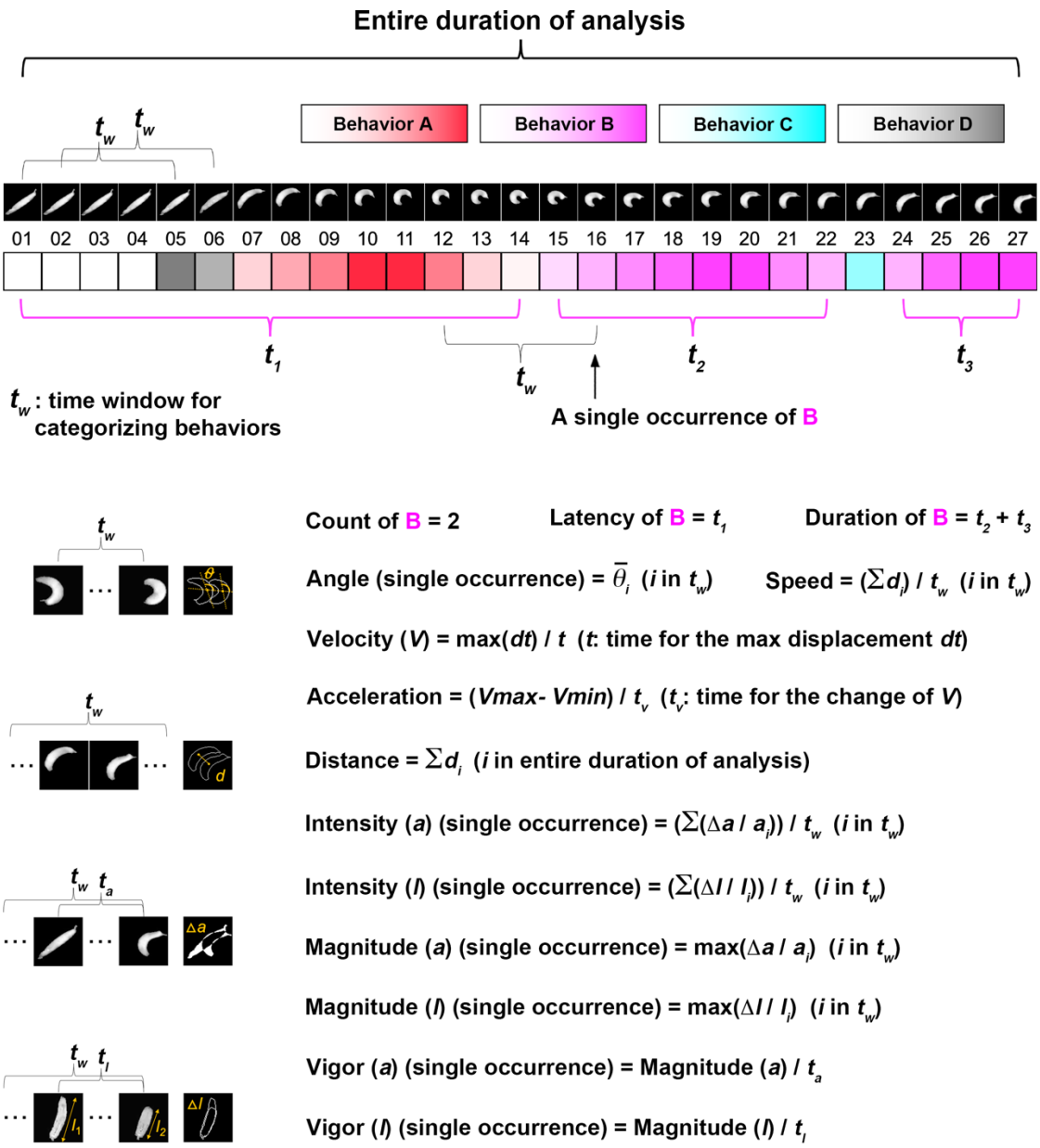

Figure S4. The Quantifier analyzes diverse aspects of a behavior, related to Figures 5 and 6.
A schematic that shows the process of frame-wise categorizations on behaviors, and how the Quantifier uses the information of behavioral categories and animal foregrounds to calculate 14 behavioral parameters in *LabGym*.

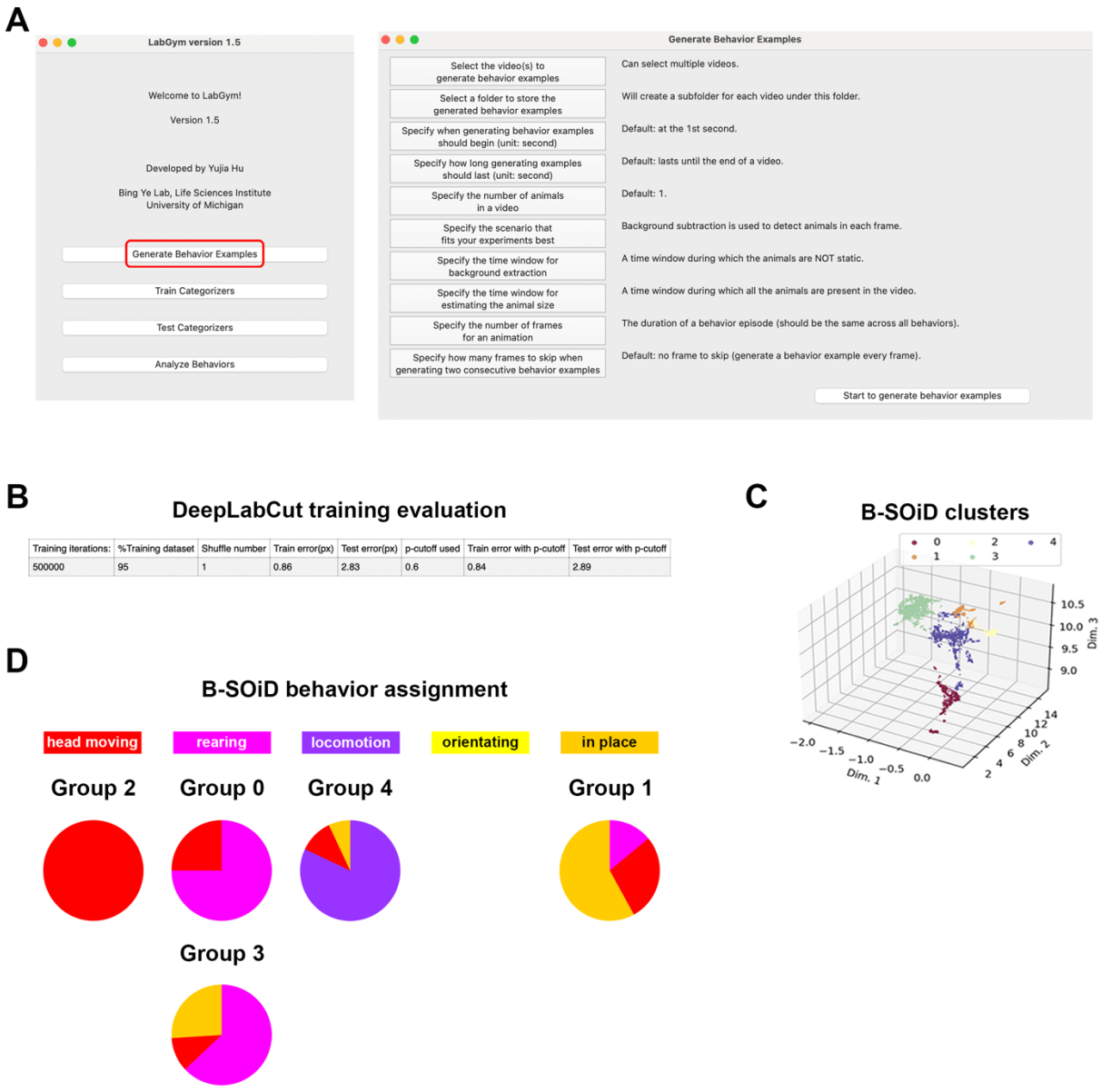

Figure S5. The benchmark comparison between DLC + B-SOiD and *LabGym*, related to Figure 7.
A. The GUI in *LabGym* demonstrated with Generate Behavior Examples functional unit.
B. The evaluation of training in DLC on the training dataset.
C. The best fit of clustering (5 groups) in B-SOiD using the outputs from DLC on the training dataset.
D. The behavior assignment of the 5 groups clustered by B-SOiD on the training dataset. All video examples in group 2 showed head movement behavior and group 2 was assigned as 'head movement'; most of the videos examples in group 0 and group 3 showed rearing behavior and group 0 and group 3 were assigned as 'rearing'; most of the video examples in group 4 showed locomotion behavior and group 4 was assigned as 'locomotion'; none of the video examples in all groups showed complete orientating behavior; most of the video examples in group 1 showed in place resting behavior and group 1 was assigned as 'in place'.

LarvaN (overall accuracy: 0.98)			
Animation Analyzer	8 x 8 x 1, level 1		
Pattern Recognizer	32 x 32 x 3, level 2		
	precision	recall	f1
crawling	0.99	0.97	0.98
curling	0.97	0.97	0.97
immobile	0.98	1	0.99
rolling	0.97	0.98	0.97
turning	0.98	0.98	0.98
uncoiling	0.96	0.95	0.96

Categorizer for Larvae #2 (overall accuracy: 0.93)			
Animation Analyzer	8 x 8 x 1, level 1		
Pattern Recognizer	32 x 32 x 3, level 3		
	precision	recall	f1
crawling	0.94	0.96	0.95
curling	0.92	0.88	0.9
immobile	0.97	0.95	0.96
rolling	0.94	0.95	0.95
turning	0.89	0.93	0.91
uncoiling	0.92	0.91	0.91

Categorizer for Rats #1 (overall accuracy: 0.74)			
Animation Analyzer	16 x 16 x 1, level 2		
Pattern Recognizer	32 x 32 x 3, level 3		
	precision	recall	f1
body grooming	0.69	0.73	0.71
face grooming	0.48	0.68	0.56
head swaying	0.66	0.74	0.7
locomotion	0.91	0.91	0.91
orientating	0.8	0.8	0.8
rearing	0.8	0.65	0.72
in place	0.58	0.56	0.57
still	0.89	0.82	0.85

Categorizer for Rats #2 (overall accuracy: 0.8)			
Animation Analyzer	32 x 32 x 1, level 4		
Pattern Recognizer	64 x 64 x 3, level 4		
	precision	recall	f1
body grooming	0.78	0.71	0.74
face grooming	0.65	0.5	0.57
head swaying	0.74	0.76	0.75
locomotion	0.95	0.95	0.95
orientating	0.88	0.91	0.89
rearing	0.8	0.85	0.83
in place	0.69	0.65	0.67
still	0.8	0.86	0.83

Categorizer for Rats #3 (overall accuracy: 0.76)			
Animation Analyzer	32 x 32 x 1, level 5		
Pattern Recognizer	64 x 64 x 3, level 5		
	precision	recall	f1
body grooming	0.8	0.6	0.69
face grooming	0.53	0.56	0.54
head swaying	0.72	0.79	0.76
locomotion	0.89	0.86	0.88
orientating	0.81	0.83	0.82
rearing	0.78	0.81	0.79
in place	0.62	0.61	0.61
still	0.84	0.84	0.84

RatA (overall accuracy: 0.88) (after refining labeling)			
Animation Analyzer	32 x 32 x 1, level 4		
Pattern Recognizer	64 x 64 x 3, level 4		
	precision	recall	f1
body grooming	0.89	0.88	0.88
face grooming	0.82	0.8	0.81
head swaying	0.75	0.88	0.81
locomotion	0.95	0.98	0.96
orientating	0.96	0.92	0.94
rearing	0.88	0.88	0.88
in place	0.87	0.7	0.77
still	0.89	0.93	0.91

Categorizer for Mice #1 (overall accuracy: 0.85)			
Animation Analyzer	32 x 32 x 1, level 2		
Pattern Recognizer	32 x 32 x 3, level 2		
	precision	recall	f1
behind the wheel	0.94	0.94	0.94
body grooming fv	0.67	0.44	0.53
body grooming sv	0.82	0.62	0.71
chewing fv	0.63	0.68	0.65
chewing sv	0.74	0.85	0.79
coming down	0.89	0.99	0.94
face grooming fv	0.66	0.83	0.74
face grooming sv	0.86	0.67	0.75
foraging fv	0.85	0.89	0.87
foraging sv	0.82	0.93	0.87
rearing up	0.84	0.83	0.83
resting on the wheel	0.91	0.92	0.92
running on the wheel	0.93	0.93	0.93
sniffing fv	0.59	0.57	0.58
sniffing sv	0.87	0.74	0.8
standing	0.93	0.87	0.9
turning front	0.71	0.75	0.73
turning side	0.78	0.67	0.72
unknown bv	0.94	0.9	0.91
walking	0.93	0.9	0.91

MouseH (overall accuracy: 0.9)			
Animation Analyzer	64 x 64 x 1, level 4		
Pattern Recognizer	64 x 64 x 3, level 4		
	precision	recall	f1
behind the wheel	0.94	0.91	0.93
body grooming fv	0.71	0.83	0.77
body grooming sv	0.85	0.79	0.82
chewing fv	0.75	0.72	0.73
chewing sv	0.89	0.78	0.83
coming down	0.96	0.96	0.96
face grooming fv	0.85	0.93	0.89
face grooming sv	0.83	0.83	0.83
foraging fv	0.93	0.92	0.93
foraging sv	0.85	0.94	0.9
rearing up	0.92	0.92	0.92
resting on the wheel	0.97	0.85	0.9
running on the wheel	0.91	0.98	0.94
sniffing fv	0.68	0.57	0.62
sniffing sv	0.84	0.89	0.86
standing	0.94	0.95	0.94
turning front	0.76	0.81	0.79
turning side	0.92	0.79	0.85
unknown bv	0.97	0.96	0.97
walking	0.88	0.92	0.9

Table S1. Example showing the training metrics for all tested Categorizers in selecting the ones that are most suitable for each of the 3 behavioral datasets, related to Figure 4.
We started from the simplest networks for each dataset and gradually increase the network complexity until the training performance was satisfying.