

# datacard

## WhACC Dataset Data Card

### Motivation

- **Purpose:** This dataset was created to identify touch periods from high-speed videos of head-fixed behaving rodents, with the goal of greatly reducing the manual review time necessary for behavioral analyses.
- **Creators:** Phillip Maire, Samson G. King, Jonathan Andrew Cheung, and Stef Walker from the University of Southern California
- **Funding:** the following grants funded this dataset 5U01NS120824 and 5R01NS102808

### Composition

- **Content:** The dataset contains instances representing high-speed video frames (1000 frames per second) capturing a moving mouse whisker contacting a circular pole. Images are cropped at 61X61 pixels.
- **Sessions:** The data were collected from eight different behavioral sessions across two different laboratories, utilizing three different experimental rigs.
- **Number of Instances:** A total of 1,919,822 images are included in this dataset across all 8 sessions, and a total 1,149,355 of these frames have the round pole 'in-range' where the whisker can contact the pole.

### Collection Process

- **Videography:** The data were collected using backlit high-speed cameras to capture the detailed movements of rodent whiskers.
- **Ethics:** All procedures were approved under USC IACUC protocols 20169 and 20788 in accordance with United States national guidelines issued by Office of Laboratory Animal Welfare of the National Institute of Health.

### Preprocessing/Cleaning/Labeling

- **Preprocessing:** Videos were processed to focus on a 61x61 pixel window surrounding the object being touched. Frames where the object was out of the reach of the whiskers were excluded from the dataset.

- **Labeling:** Touch labels for frames were assigned by three expert human curators and contacts were given by majority rules

## Uses

- **Intended Use:** The dataset was specifically collected to develop and train machine learning models, such as WhACC (<https://github.com/hireslab/whacc>), to determine the touch times of whiskers on objects during behavioral assessments. This is particularly relevant for neuroscience research encompassing sensory encoding and sensorimotor integration.

## Distribution

- **Accessibility:** The dataset is publicly available and can be accessed at <https://drive.google.com/drive/folders/1nErTbg2TTaZawznJalc1EGKOQHhy55C1> or upon request if the link is inaccessible.

## Maintenance

- **Hosting:** The dataset will be hosted via a laboratory Dropbox under the supervision of Dr. Andrew Hires' lab ([hireslab.org](http://hireslab.org)).
- **Updates:** Any updates to the dataset, including corrections or additions, will be communicated through the readme on the WhACC (<https://github.com/hireslab/whacc>) GitHub and reflected in the hosted dataset accordingly.