MindSet: Vision. A toolbox for testing DNNs on key psychological experiments

Valerio Biscione[1][*]   **Dong Yin**[1]   **Gaurav Malhotra**[2]   **Marin Dujmovic**[3]
Milton L. Montero[4]   **Guillermo Puebla**[5]   **Federico Adolfi**[1,6]
Rachel F. Heaton[7,8]   **John E. Hummel**[8]   **Benjamin D. Evans**[9]   **Karim Habashy**[1]
Jeffrey S. Bowers[1]

[1]School of Psychological Science, University of Bristol
[2]University at Albany, State University of New York
[3]School of Physiology, Pharmacology and Neuroscience, University of Bristol
[4]IT University of Copenhagen
[5]Instituto de Alta Investigación, Universidad de Tarapacá, Chile
[6]Ernst Strüngmann Institute for Neuroscience, Max-Planck Society, Germany
[7]Siebel Center for Design, University of Illinois Urbana-Champaign
[8]Department of Psychology, University of Illinois Urbana-Champaign
[9]School of Engineering and Informatics, University of Sussex

1 Supplementary Materials: Accessibility and Licensing

Pre-generated datasets are accessible for viewing and download on Kaggle in both a comprehensive version and a "lite" version, which can be found at large and "lite" respectively. `Croissant` metadata for both versions are automatically generated by Kaggle and can be downloaded from the Kaggle page. These datasets comprise a collection of folders containing `.png` images and associated metadata. Notably, each dataset's root folder includes a `.csv` file containing metadata for each individual image, as well as a `.toml` file detailing the configuration parameters used in dataset generation. This structure enables precise replication of each dataset published on Kaggle using our provided codebase. The dataset is made available under the CC 1.0 Universal license, with a DOI of `10.34740/kaggle/dsv/8037793`.

The codebase for regenerating the dataset with varying parameters is hosted on GitHub, licensed under MIT. Comprehensive documentation on dataset generation procedures is included in the main README file within the repository. Additionally, we provide code for evaluating the datasets using three distinct testing methods: out-of-distribution classification, similarity analysis, and decoder approach. Each method is documented in a dedicated README file within the corresponding folder, such as `src/utils/similarity_judgment/REAMDE.md`, and examples are provided. The appendix in the main text offers detailed descriptions of each dataset, emphasizing that each dataset can be regenerated by specifying a set of parameters. The meaning of each parameter for each dataset is accessible via the `-h` flag and is also provided in the accompanying `.toml` file used for dataset generation. The entire dataset generation process is tested on Windows Server 2022 (Windows 10/11), macOS 14.4.1, and Ubuntu 20.04. Authors assume full responsibility in the event of any rights violations.

---

[*]corresponding author: valerio.biscione@gmail.com