MDMC
Master in Data Management and Curation

AREA SCIENCE PARK

SISSA

Sensorium Project

Applied AI course

Students: Flo Ana, Longato Enrico, Nana Njantang Epse Osazuwa Ruth, Nkana Ngan Valentin, Nyandu Kagarabi Emmanuel, Palacios Flores Luis Fernando, Talukdar Smritirekha, Tsoptio Fougang Lesly

Lecturers: Matteo Biagetti, Emanuele Panizon, and Tommaso Rodani

Laboratory of Data Engineering-Area Science Park (Trieste), 5th December 2025

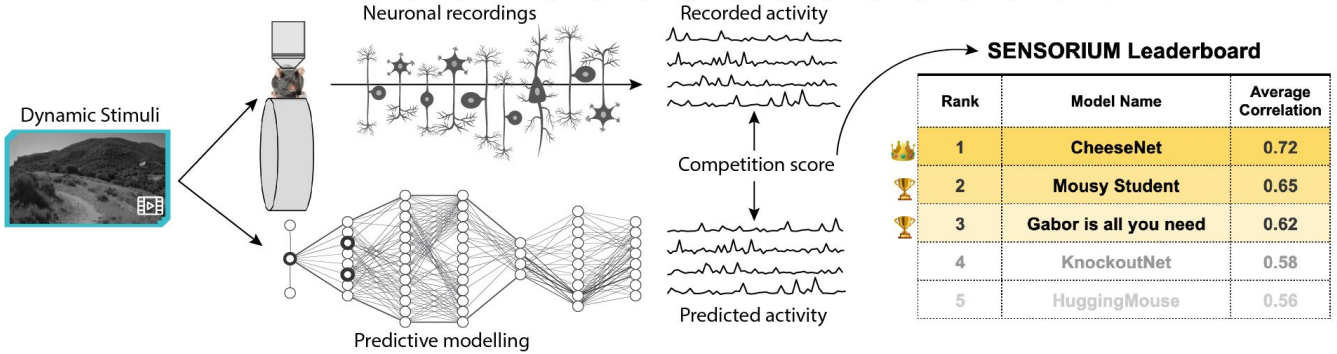
Outline

- ➔ Introduction
- ➔ Dataset description
- ➔ Workflow plan
- ➔ Sensorium webapp
- ➔ Conclusion and perspectives

Introduction

- The **Sensorium2023** dataset provides ***neural***, ***behavioural***, and **stimulus data** recorded from **five mice** during visual processing experiments.
  - Each dataset is delivered as a `.zip` file containing two folders:
    - **Data**, which holds trial-wise NumPy arrays for presented videos, neural responses, pupil position, and behaviour;
    - and **meta**, which includes neuron-level information, experimental **statistics**, and **trial** metadata.
- Trials are stored in **randomized order**, and test-set neural responses are withheld for competition scoring.
- For a full technical description, refer to the accompanying white paper.

Dataset Description



SENSORIUM Leaderboard

Rank	Model Name	Average Correlation
1	CheeseNet	0.72
2	Mousy Student	0.65
3	Gabor is all you need	0.62
4	KnockoutNet	0.58
5	HuggingMouse	0.56

what do we have?

Broad dataset:

Experimental variables in NumPyformat

Data

- *Videos, responses, behaviour, pupil centre*

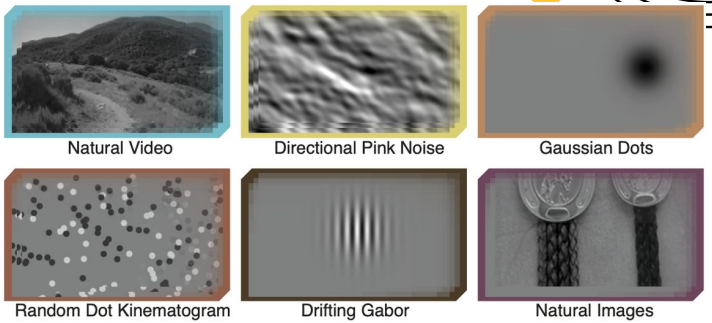
Meta

- *Neuron coordinates*
- *Statistics of experimental variables*

Dataset : [SENSORIUM 2023 | Welcome to the Sensorium 2023 Competition!]

Additional info

Video Type	Duration	Not NaN frames	Particularity
Natural Movies	-	-	-
Gabor	10 sec	300	each last 833 ms (~25 frames)
Directional Pink Noise	10.8 sec	324	each last 900 ms (~27 frames)
Random Dot	8 sec	240	each movie lasts 2 s (~60 frames)
Natural Images	10 sec	300	15 frames are repeated preceded by gray screen lasting 400-600 ms from 12 to 18 frames
Gaussian Dot	10.5 sec	315	9 frames are repeated

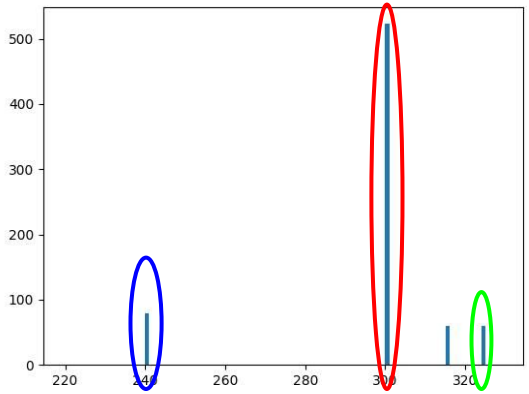


All videos have dimension 36x64 for 324 frames at a presentation rate of 30Hz

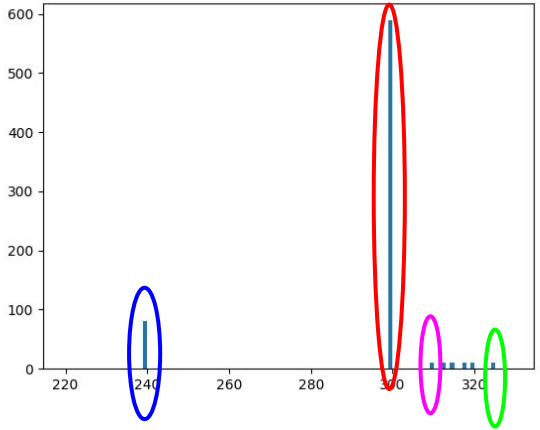
If a video has less than 324 frames, the remaining frames are white.

Distribution of not white frames

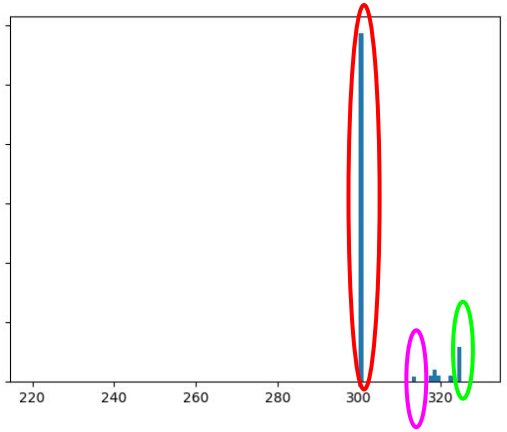
mouse 11-10



mouse 6-9



mouse 2-10



Random Dot

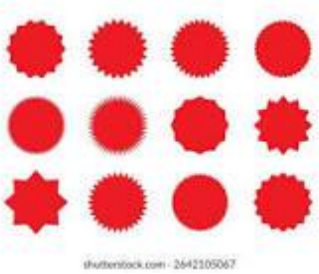
Gabor, Natural Images

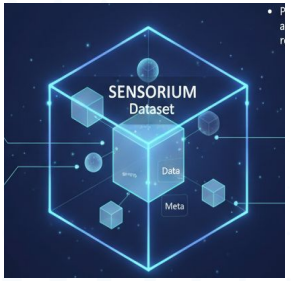
Gaussian Dot

Directional pink noise

Natural Movies

Workflow plan





Label videos by types naturalistic, gaussian, waves…

Combine videos and neural responses with all relevant metadata

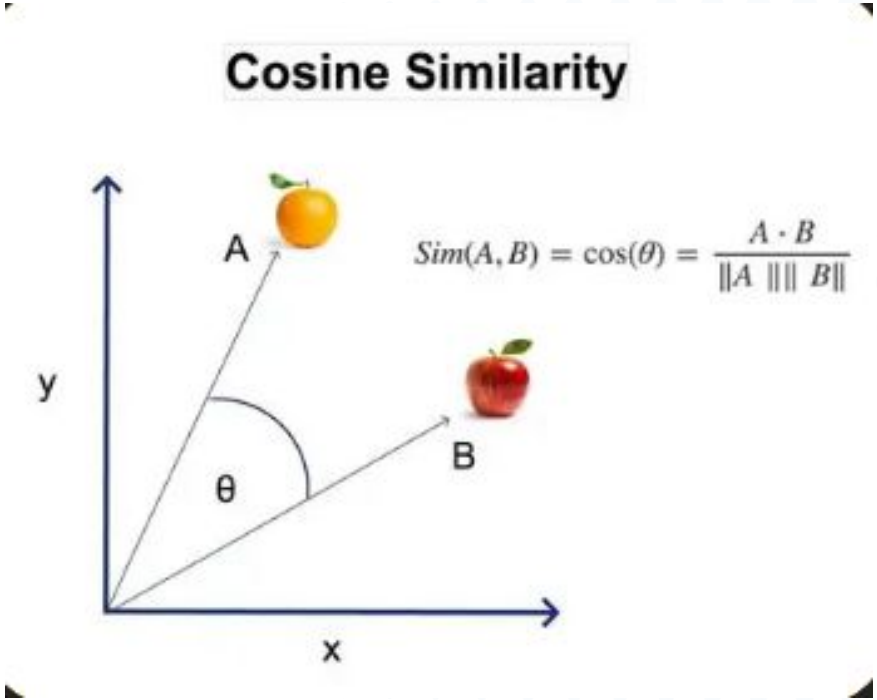
Build visualisation tool for exploration of the whole dataset

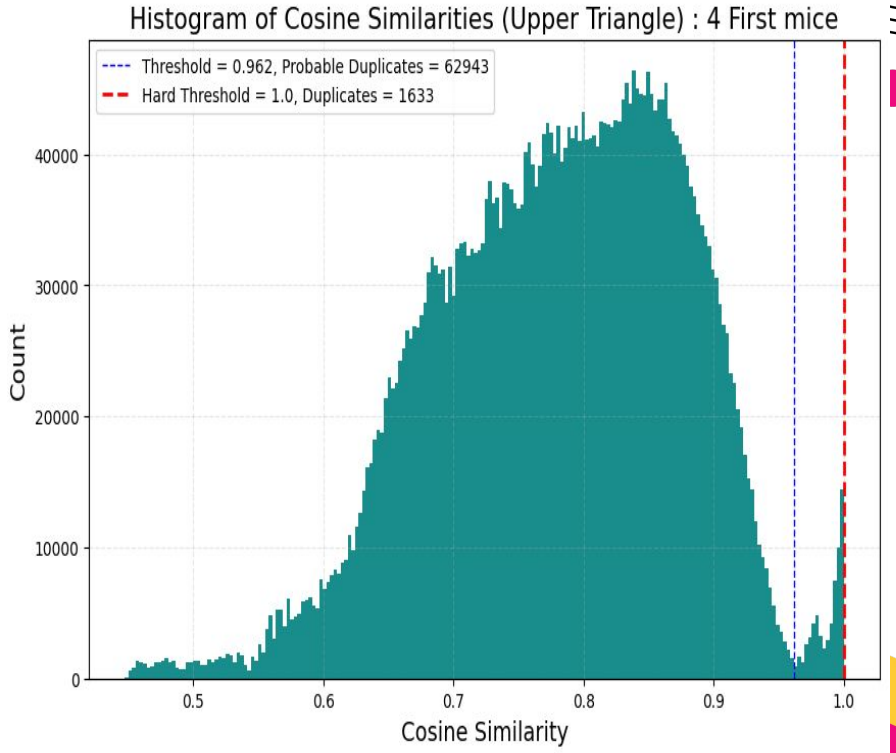
Created a unique label

	recording	file	label	valid_frames	recording_short	unique_label
0	dynamic29228-2-10-Video-8744edeac3b4d1ce16b680...	407.npy	NaturalVideo	300	2-10	2-10_407_NaturalVideo
1	dynamic29228-2-10-Video-8744edeac3b4d1ce16b680...	379.npy	NaturalVideo	300	2-10	2-10_379_NaturalVideo
2	dynamic29228-2-10-Video-8744edeac3b4d1ce16b680...	18.npy	NaturalVideo	300	2-10	2-10_18_NaturalVideo
3	dynamic29228-2-10-Video-8744edeac3b4d1ce16b680...	270.npy	NaturalVideo	300	2-10	2-10_270_NaturalVideo
4	dynamic29228-2-10-Video-8744edeac3b4d1ce16b680...	241.npy	NaturalVideo	300	2-10	2-10_241_NaturalVideo
...	...	...	...	...	...	...
3585	dynamic29514-2-9-Video-8744edeac3b4d1ce16b6809...	478.npy	NaturalVideo	300	2-9	2-9_478_NaturalVideo
3586	dynamic29514-2-9-Video-8744edeac3b4d1ce16b6809...	418.npy	NaturalVideo	300	2-9	2-9_418_NaturalVideo
3587	dynamic29514-2-9-Video-8744edeac3b4d1ce16b6809...	97.npy	Gabor	300	2-9	2-9_97_Gabor
3588	dynamic29514-2-9-Video-8744edeac3b4d1ce16b6809...	210.npy	NaturalVideo	300	2-9	2-9_210_NaturalVideo
3589	dynamic29514-2-9-Video-8744edeac3b4d1ce16b6809...	131.npy	NaturalVideo	300	2-9	2-9_131_NaturalVideo

3590 rows × 6 columns

Cosine Similarities on Embeddings





Threshold = 1.00

Reference : Mouse_1__11_10__101.npy

Duplicate : Mouse_1__11_10__210.npy

Unified JSON Structure

Advantages

- ➔ Faster, More Flexible Data Access
- ➔ Facilitates Exploratory Analysis and Visualization
- ➔ Improves Reproducibility
- ➔ Scales to Large Datasets

Tools

Hardware Tools

Software Tools

git

Link to repository

Sensorium Web App

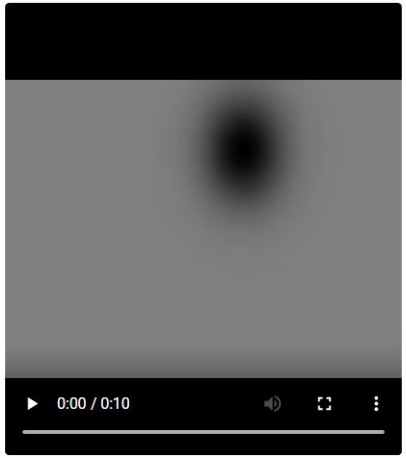
User Interface (UI) Overview

MOUSE ID DROPDOWN: dynamic29156-11-10-Video-8744edeac3b4d1ce16b680916b5267ce

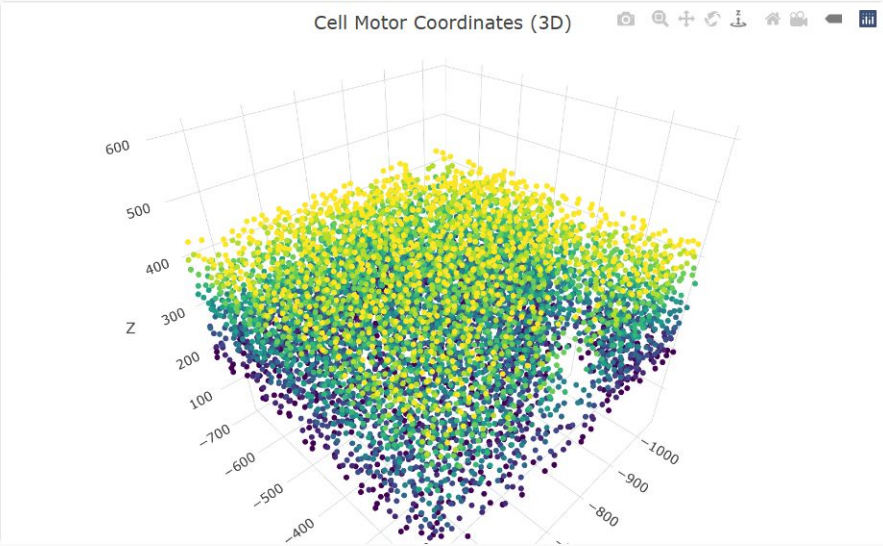
REPRESENTATIVE VIDEO DROPDOWN:

12

GaussianDot



CELL MOTOR COORDINATES



ADDITIONAL METADATA

Video ID: 12

Total Valid Frames: 315

Number of Equivalent Videos: 9

Consistent Responses: ✗ No

Incorrect Responses on Valid Frames: 9

Equivalent Video IDs: 14, 23, 28, 57, 155, 316, 409, 590, 726

Conclusion and perspectives

Conclusion

- ➔ *Labelled videos by types (naturalistic, Gaussian, waves) and eventually subtypes*
- ➔ *combine videos and neural responses with all relevant metadata in a data structure that allows an efficient query for preliminary exploration*
- ➔ *Built visualisation tool for exploration of the whole dataset*
- ➔ *Developed of a web application using **Django***

Perspectives

- ➔ *Deep learning for predictive modelling*
- ➔ *Improving the Web app*

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

nffa-di E-ARGO beanTech IT moves your business CERIC

Thank you!

Grazie!

This Pilot training activity has been funded by the European Union – NextGenerationEU within the PNRR projects funded pursuant to Article 11, paragraph 1, of Notice 594/2024:

- "NFFA-DI cod. IR0000015, Missione 4, "Istruzione e Ricerca" – Componente 2, "Dalla ricerca all'impresa" – Linea di investimento 3.1, "Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione" – Azione 3.1.1, "Creazione di nuove IR o potenziamento di quelle esistenti che concorrono agli obiettivi di Eccellenza Scientifica di Horizon Europe e costituzione di reti" (CUP B53C22004310006).
- "EFC cod. SSU2024-00002, Missione 4 "Istruzione e ricerca" - Componente 1, "Potenziamento dell'offerta dei servizi all'istruzione: dagli asili nido all'universita'" - Investimento 3.4 "Didattica e competenze universitarie avanzate" - Sub-Investimento "Rafforzamento delle scuole universitarie superiori" (CUP: G97G24000100007).