University of Milano – Bicocca
Master's Degree in Data Science
Academic Year 2022-2023

Brain-to-image: Image classification and retrieval based on EEG signals

Author: Gianluca SCURI

Supervisor: Prof. Paolo NAPOLETANO
Co-supervisor: Prof. Simone BIANCO

Decoding human brain

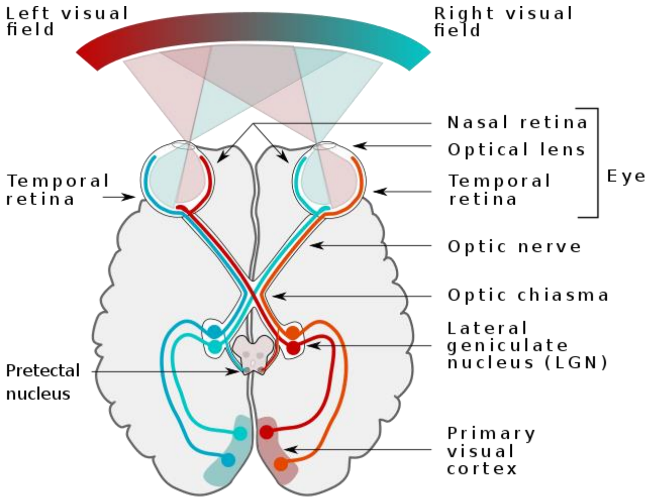
- Electroencephalography (EEG) to study **visual perception**
- Project **aims:**
  - **Classify EEG signals** based on image classes
  - **Reconstruct** the original image

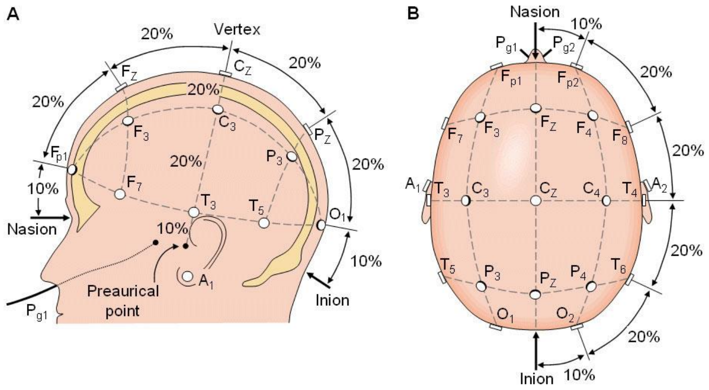
- Potential **use cases:**
  - Assess and treat **visual dysfunction**
  - Assist patients with **communication difficulties**
  - Enhance experiences in **VR/AR**

- Neurobiology, Neuroscience and Artificial Intelligence

Human brain and visual perception

- ❑ 86 billions of interconnected neurons
- ❑ **Brain waves:**
    - ▪ Delta (0.5-4Hz)
    - ▪ Theta (4-8Hz)
    - ▪ Alpha (8-12Hz)
    - ▪ Beta (12-35Hz)
    - ▪ Gamma (>35Hz)
- ❑ **EEG recordings**
- ❑ 2 hemisphere, **4 lobes**
- ❑ Eyes → Optic nerve → Occipital lobe

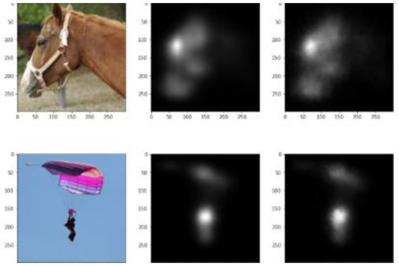




Related works

- *Spampinato et al. 2017-2020*:
    - **83% accuracy** (40 classes)
    - Image reconstruction with VAE and GAN
- *Li et al. 2021*:
    - Proved that the **dataset is flawed**
- *Gifford et al. 2022*:
    - Released a **new large dataset**
    - Two papers released on this dataset *Du et al.* & *Song et al.* mid 2023

Khare et al. 2022

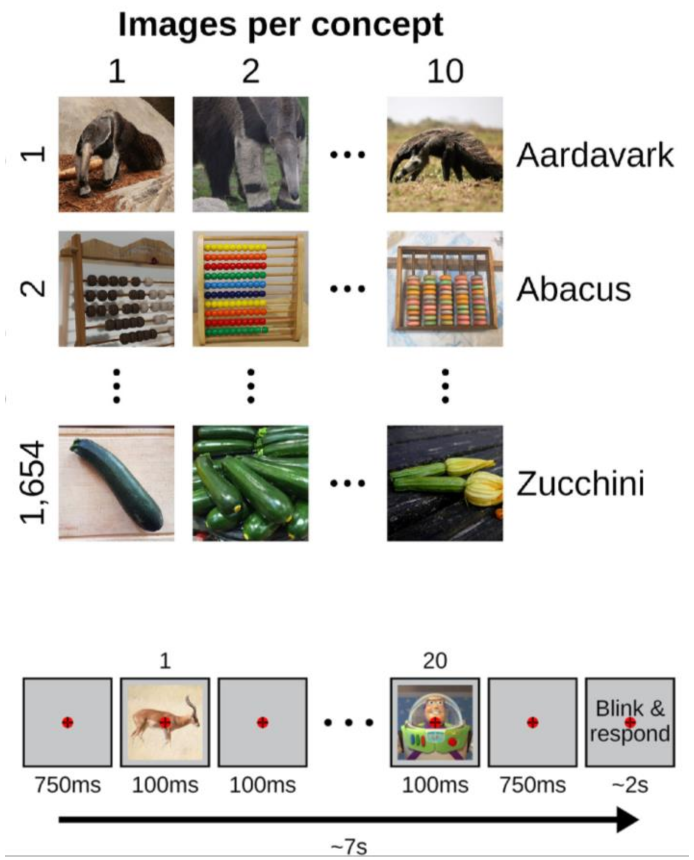


Khaleghi et al. 2022.

Singh et al. 2023

Dataset: THINGS-EEG[1]

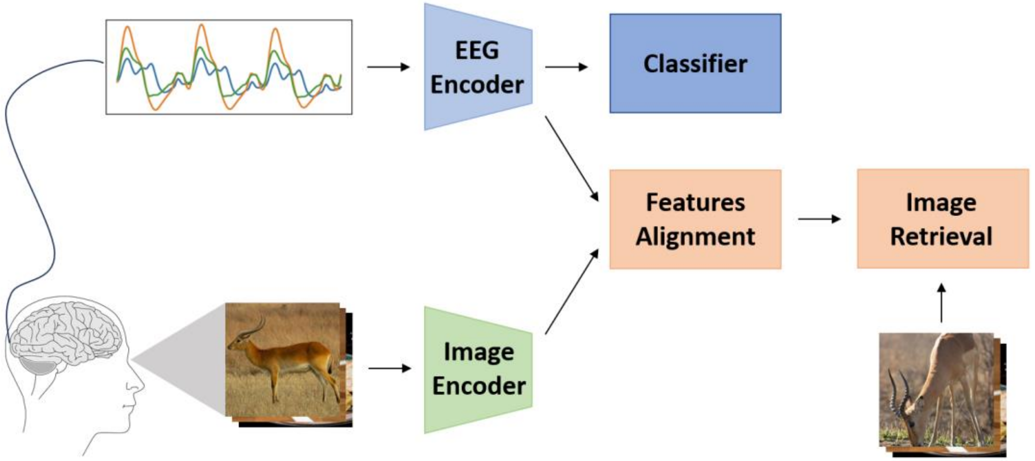
- **THINGS project**
- Rapid Serial Visual Presentation (**RSVP**) paradigm
- Specs:
  - 20 minutes (100ms per image)
  - **1654 concepts** x 10 images x 4 rep. = **66160 trials**
  - **27 high-level categories**



[1] Alessandro T Gifford et al. "A large and rich EEG dataset for modelling human visual object recognition". NeuroImage 2022

Project framework

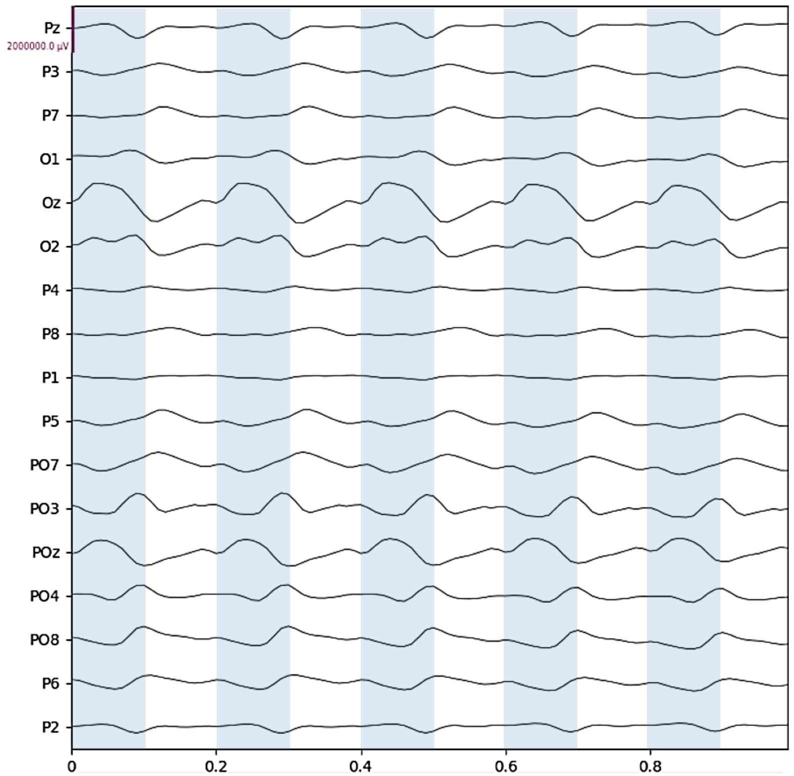
- **EEG classification**
  - By concepts (SLMC)
  - By categories (MLMC)
- **Image retrieval**
  - Features alignment
  - Image retrieval

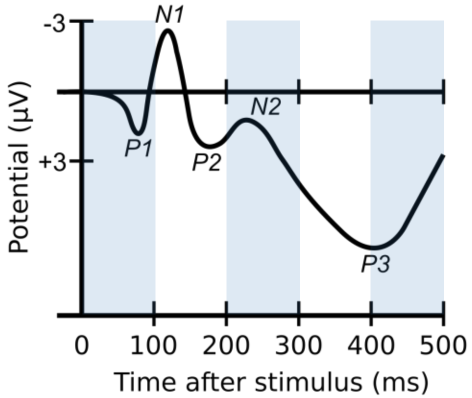


- **Key concepts:**
  - **EEG signals contain information** about the image
  - EEG **features can be extracted** from this signals
  - **Features can be used** for classification or retrieval

EEG preprocessing

- **Epoching**
  - Original: 1000ms windows
  - Proposed: 500ms windows
- **Subsampling**
  - Original: 100Hz
  - Proposed: 200Hz
- **Channel selection** (17 ch. on occipital and parietal lobes)
- **Baseline correction** (mean value in range 0-200ms)
- Multivariate Noise Normalization (**MVNN**)





EEG models

- ❑ LSTM and CONV1D

Layers
LSTM
Avg. Pool 1D
Conv. 1D
Avg. Pool 1D
Flatten&Linear

- ❑ Temporal and Spatial CONV2D

Layers
Conv. 2D
Avg. Pool 2D
Conv. 2D
Flatten&Linear

- ❑ Spectrogram CONV2D

Layers
Conv. 2D
Avg. Pool 2D
Conv. 2D
Avg. Pool 2D
Flatten&Linear

- ❑ EEGNet [1]

Layers
Conv. 2D
Conv. 2D
Avg. Pool 2D
Conv. 2D
Conv. 2D
Avg. Pool 2D
Flatten&Linear

- ❑ All models are based on **neural networks**
- ❑ Activation functions, batch normalization and dropout are also used

[1] Vernon J Lawhern et al. “EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces”. Journal of neural engineering 2018

EEG classification

Classification by concepts: results

- Model comparison:
  - LSTM model obtain **15x higher than chance** accuracy (0.06%)
  - LSTM model performed better than the EEGNet
  - Limits: EEG, RSVP, high number of classes

	Top-1 accuracy
LSTM	**0.94**
TSConv	0.81
SpectrConv	0.16
EEGNet	0.62

- EEG **preprocessing comparison** (top-1 accuracy on 100 concepts):
  - Up to **60% improvements** with proposed method

	Original preproc.	Proposed preproc.
LSTM	5.64	**7.64**
TSConv	5.56	**9.20**
SpectrConv	**2.43**	1.99
EEGNet	7.38	**9.11**

Classification by categories: results

- Fairly **balanced recall and precision** thanks to concept weighting and thresholds
- Limits
  - **Reduced time** for the brain to process such a high semantic information
  - High **intra-class heterogeneity** (tools, parts of car, …)

	F1	Precision	Recall	Accuracy
LSTM	3.66	2.92	4.97	20.33
TSConv	8.78	7.51	**11.01**	19.43
SpectrConv	5.68	4.71	7.35	19.07
EEGNet	**9.37**	**8.31**	10.75	**20.82**

Image retrieval based on EEG

Image retrieval: features alignment

- Features extraction:
  - Semantic similarity vs visual similarity
  - **EEG encoder:** pre-trained EEGNet on 100 classes
  - **Image encoder**: pre-trained ResNet18
  - 100 components each (PCA)

- Features alignment:
  - Regression with linear model
  - MSE loss and CLIP[1] loss

N. components	Top-1 Accuracy
512	67.48
200	66.34
100	63.64
50	58.19

[1] Alec Radford et al. “Learning transferable visual models from natural language supervision”. International conference on machine learning 2021

Image retrieval: queryable dataset

- **Comparison** with generative models:
  - Limited to existing images only
  - No visual artifacts, easier to train and computational less expensive
- **Dataset structure:**
  - KDTree with euclidean metric
  - 1535 images from THINGS dataset (100 classes)
- **Queries:**
  - 10 nearest neighbors

Image retrieval: results

- ❑ Results **higher than chance**
- ❑ CLIP loss function performed better than MSE loss
- ❑ Limited by EEG encoder

	Same image accuracy	Same class accuracy
MSE loss	0.81	10.02
CLIP loss	**2.13**	**14.11**
Chance	0.65	9.38

Conclusions

- **Classification:**
  - EEG signals collected with **RSVP** can be classified
  - EEG contain information from a **high number of classes**
  - **Improved original preprocessing** and designed a better model than EEGNet
  - Results limited by the **rigid approach**

- **Image retrieval:**
  - Can be a **valid approach** to assist patients
  - Approach limited by **EEG encoder**

Future works

- Deepen classification by exploiting **multimodal approaches:**
    - Textual information for concepts and categories

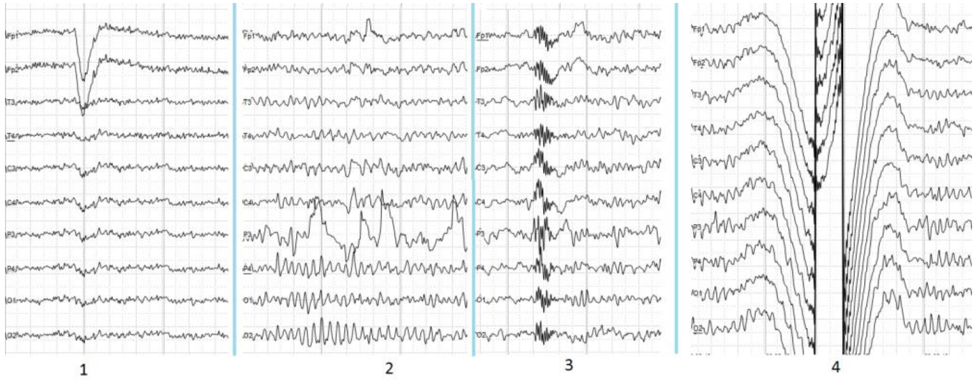
- **Improve EEG encoder:**
    - **Data augmentation**
    - 3D arrangement of electrodes (**graph** convolutional networks)
    - Using **autoencoder**

- Deepen **visual similarity**:
    - **Saliency** maps
    - **Masking** or **segmentation**

Thank you for your attention

EEG strengths and weaknesses

- **Strengths:**
  - Non-invasive
  - High temporal resolution
  - Cheap, compact and silent
- **Weaknesses:**
  - Electrode placement takes time
  - Low spatial resolution (blurring effect)
  - Low signal-to-noise ratio
  - Signals can vary greatly from person to person and even from session to session



1. Eyes movement
2. Bad contact electrodes
3. Swallowing
4. Bad contact reference electrode

EEG Models

❑ LSTM + CONV1D

Layer	In	Out	Kernel	Stride	Dimension
LSTM	C	K1			(b,K1,T)
Avg. Pool 1D	K1	K1	M1	S1	(b,K1,T/S1)
Conv. 1D	K1	K2	M2	S2	(b,K1,T/S1)
Avg. Pool 1D	K2	K2	M3	S3	(b,K1,T/S1*S3)
Flatten&Linear					(b, n_classes)

❑ Temporal and Spatial CONV2D

Layer	In	Out	Kernel	Stride	Dimension
Conv. 2D	1	K	(1,M1)	(1,S1)	(b,K,C,T-m1+1)
Avg. Pool 2D	K	K	(1,M2)	(1,S2)	(b,K,C,(T-M1-M2+2)/S2)
Conv. 2D	K	K	(C,1)	(1,1)	(b,K,1,(T-M1-M2+2)/S2)
Flatten&Linear					(b, n_classes)

❑ Spectrogram CONV2D

Layer	In	Out	Kernel	Stride	Dimension
Conv. 2D	C	K1	(M1,M1)	(1,1)	(b,K1,N,M)
Avg. Pool 2D	K1	K1	(M2,M2)	(S1,S1)	(b,K1,N/S1,M/S1)
Conv. 2D	K1	K2	(M3,M3)	(1,1)	(b,K1,N/S1,M/S1)
Avg. Pool 2D	K2	K2	(N/S1,N/S1)	(1,1)	(b,K2,1,1)
Flatten&Linear					(b, n_classes)

❑ EEGNet

Layer	In	Out	Kernel	Stride	Dimension
Conv. 2D	1	K1	(1,M1)	(1,1)	(b,K1,C,T)
Conv. 2D	K1	K2	(C,1)	(1,1)	(b,K2,C,T)
Avg. Pool 2D	K2	K2	(1,M2)	(1,S1)	(b,K2,C,(T-M2+1)/S1)
Conv. 2D	K2	K2	(1,M3)	(1,1)	(b,K2,C,(T-M2+1)/S1)
Conv. 2D	K2	K3	(1,1)	(1,1)	(b,K3,C,(T-M2+1)/S1)
Avg. Pool 2D	K3	K3	(1,M2)	(1,S1)	(b,K3,C,(T-M2+1)/(S1*S1))
Flatten&Linear					(b, n_classes)

Classification by concept

- Different subsets comparison

	20 classes	50 classes	100 classes	200 classes	1654 classes
LSTM	16.14	11.46	7.64	4.47	**0.94**
TSConv	**17.19**	10.59	**9.20**	4.65	0.81
SpectrConv	14.58	5.52	1.99	1.39	0.16
EEGNet	14.58	**11.81**	9.11	**5.08**	0.62

Classification by categories

	F1	Precision	Recall	N concepts
animal	32.3	36.3	29.2	177
bird	7.3	6.8	7.9	27
body part	19.1	11.7	52.0	34
clothing	22.5	18.2	29.5	108
clothing accessory	5.3	4.1	7.6	38
container	6.9	11.9	4.9	105
dessert	4.8	5.0	4.6	37
drink	3.2	3.5	3.0	19
electronic device	10.0	8.5	12.4	74
food	28.9	38.4	23.1	295
fruit	7.3	5.6	10.3	34
furniture	7.9	5.6	13.6	39
home decor	8.2	7.6	8.9	45
insect	1.7	3.0	1.1	17
kitchen appliance	2.7	2.6	2.8	20
kitchen tool	NA	0.0	0.0	27
medical equipment	5.6	4.0	9.3	27
musical instrument	2.8	3.6	2.2	33
office supply	6.5	4.3	13.1	25
part of car	4.4	3.3	6.5	30
plant	12.9	8.7	25.0	47
sports equipment	4.0	5.6	3.1	64
tool	4.6	6.5	3.6	107
toy	3.6	5.4	2.7	34
vegetable	6.2	4.7	8.9	42
vehicle	7.7	7.8	7.6	70
weapon	1.1	3.8	0.7	48

RSVP example

WELCOME TO THIS EXPERIMENT

Sequences of images will be presented to you.

Your task is to report whether BUZZ LIGHTYEAR is present in each sequence.

If you see BUZZ, press the RIGHT ARROW key during the response period.

If you don't see BUZZ, press the LEFT ARROW key during the response period.

Be as accurate as possible.

Press any key to continue