Brain-Inspired Multimodal Deep Learning

21/11/2023

Rufin VANRULLLEN

CerCo, CNRS
PLAN: overview of past and future work

2019-2023: Chair “Deep Learning with semantic, cognitive and biological constraints”
With co-chairs C. Braud (IRIT), L. Reddy (CerCo), F. Filbet & G. Faye (IMT)

1. brain decoding with deep generative networks

2. brain-inspired recurrent networks with Predictive Coding
   • more efficient, more robust
   • more aligned with human vision (illusions)
   • detailed mathematical characterization (convergence, oscillations…)

3. brain-inspired multimodal systems based on Global Workspace (ERC GLOW 2023-2028)

2024-2028? Synergy Chair C3-PO: “Cobots with Conversation, Cognition and Perception”
Chairs: N. Asher (IRIT), T. Serre (Brown U.), O. Stasse (LAAS), R. VanRullen (CerCo)
1. Brain decoding with deep generative networks

- Mozafari, Reddy & VanRullen (2020) *IJCNN’20*
- VanRullen & Reddy (2019) *Communications Biology*
- Ozcelik et al (2022) *IJCNN’22*
- Ozcelik & VanRullen (2023) *Scientific Reports*

**images seen by subject in MRI**

**reconstructed images from fMRI signals**
1. Brain decoding with deep generative networks

Brain-Diffuser

Ozcelik & VanRullen (2023) *Scientific Reports*

**Training Stage**

- $Y_m$
- $C_{im}$
- $C_{tx}$
- CLIP-Vision
- CLIP-Text
- Regressor
- $X_{train}$
- $Y_{tx}$
- A dog licking its lips with a cupcake with candies in front of it

**Testing Stage**

- $Y_{low}$
- $\hat{C}_{im}$
- $\hat{C}_{tx}$
- Forward Diffusion
- Reverse Diffusion
- AutoKL Encoder
- AutoKL Decoder
- $\hat{Y}_{test}$
- $Y_{test}$

**Images seen**

- Dog licking lips
- Cupcake with candies

**Reconstructed images**

- Airplane
- Cityscape
- Person eating

Brain-inspired multimodal deep learning

21/11/2023
2. Predictive coding

\[
e_n(t + 1) = \beta [W_{n-1,n}^f e_{n-1}(t + 1)]_+ + \lambda d_n(t) + (1 - \beta - \lambda)e_n(t) - \alpha \nabla e_{n-1}(t)
\]

\[
d_n(t) = [W_{n+1,n}^b e_{n+1}(t)]_+
\]

\[
e_{n-1}(t) = \|e_{n-1}(t) - d_{n-1}(t)\|_2^2
\]
2. Predictive coding

- **Predify** software (for PyTorch):
  
  https://github.com/miladmozafari/predify

- We **predified**: VGG16, EfficientNetB0, ResNet18...

\[
e_n(t + 1) = \beta [W^{f}_{n-1,n} e_{n-1}(t + 1)]_+ + \lambda d_n(t) + (1 - \beta - \lambda) e_n(t) - \alpha \nabla e_{n-1}(t)
\]

\[
d_n(t) = [W^{b}_{n+1,n} e_{n+1}(t)]_+
\]

\[
e_{n-1}(t) = \|e_{n-1}(t) - d_{n-1}(t)\|_2^2
\]
2. Predictive coding

ImageNet accuracy:

- Robustness to Gaussian noise and other corruptions (ImageNet-C)
- Robustness to adversarial attacks


2. Predictive coding

• “Perception” of illusory contours, just like humans

• Mathematical analysis of convergence:
Multi-modal systems are getting very good: CLIP, DALL-E, GATO, Flamingo, PALM-E, etc.

- Vision + Language + Action + …
- Require enormous models, gigantic supervised training sets…
- How does the brain do it?
- Global Workspace Theory
3. Multimodal systems

Global Workspace (simplified) architecture

• Training objectives:
  • Supervised: N matching pairs \{\text{vis}_i, \text{lan}_i\} — we vary N to assess the need for labels
  • Unsupervised: cycle-consistency (a.k.a. « broadcast »)
    \(\Rightarrow\) only needs unpaired data!
3. Multimodal systems

Global Workspace (simplified) architecture


Brain-inspired multimodal deep learning

21/11/2023
3. Multimodal systems

Global Workspace: where we are now...

A bright red diamond, pointing to the bottom right
A green triangle at the top, towards left
An egg-shaped oval, dark-blue, pointing up, on the bottom-left of the image

Attributes = [1,0,0,-1,-0.5,-60,2,1,0,0]

B. Devillers

L. Maytie

sim2real (with LAAS)
Global Workspace: where we are going...

**ERC Advanced Project GLOW (2023-2028)**
- Develop brain-inspired multimodal deep learning systems
- Evaluate their use and relevance for machine learning
- Advance our knowledge of the brain

**ANITI Synergy Chair C3-PO (2024-2028)**
- Cobots with Conversation, Cognition & Perception
- Chairs: R. Van Rullen (CerCo), N. Asher (IRIT), T. Serre (Brown), O. Stasse (LAAS)
- Frugal multimodal robotic systems with grounded perception, language and action
- Main industrial partners: Airbus, Linagora