11-877 Advanced Multimodal Machine Learning **Spring 2022**

Week 9: Brain and Multimodal Perception

Instructors: L.-P. Morency, A. Zadeh, P. Liang *Synopsis Leads: Chonghan Chen, Martin Q. Ma*

Edited by Paul Liang *Scribes: Zhe Chen, Alex Wilf*

Follow the rest of the class here: https://cmu-multicomp-lab.github.io/adv-mmml-course/spring2022/

Summary: Multimodal machine learning is the study of computer algorithms that learn and improve through the use and experience of multimodal data. It presents unique challenges for both computational and theoretical research given the heterogeneity of various data sources.

In the discussion session of week 9, the class aimed to first exchange insights about the brain, including understanding the architecture of the brain, the intrinsic multimodal properties of the brain, mental imagery, and neural signals. The class then discussed the differences between the brain and the qualities of current AI systems, and brainstormed possible directions towards brain-inspired AI models, including some evaluation methods to compare models and the brain. The following was a list of research probes provided:

1. What are the main takeaways from neuroscience regarding unimodal and multimodal processing, integration, alignment, translation, and co-learning?
2. How can these insights inform our design of multimodal models, following the topics we covered previously (cross-modal interactions, co-learning, pretraining, reasoning)?
3. To what extent should we design AI models with the explicit goal to mirror human perception and reasoning, versus relying on large-scale pretraining methods and general neural network models?
4. What different paradigms for multimodal perception and learning could be better aligned with how the brain processes multiple heterogeneous modalities?
5. How does the human brain represent different modalities (visual, acoustic)? Are these different modalities represented in very heterogeneous ways? How is information linked between modalities?
6. What are several challenges and opportunities in multimodal learning from high-resolution signals such as fMRI and MEG/EEG?
7. What are some ways in which multimodal learning can help in the future analysis of data collected in neuroscience?

As background, students read the following papers:

1. (Required) Multimodal Images in the Brain [Kosslyn et al., 2010]
2. (Required) Multimodal Mental Imagery [Nanay, 2018]
3. (Suggested) Crossmodal Processing in the Human Brain: Insights from Functional Neuroimaging Studies [Calvert, 2001]
4. (Suggested) Deep Sparse Coding for Invariant Multimodal Halle Berry Neurons [Kim et al., 2018]
5. (Suggested) A Theoretical Computer Science Perspective on Consciousness [Blum and Blum, 2021]
6. (Suggested) Inducing Brain-relevant Bias in Natural Language Processing Models [Schwartz et al., 2019]
7. (Suggested) The Brain-IHM Dataset: a New Resource for Studying the Brain Basis of Human-Human and Human-Machine Conversations [Ochs et al., 2020]
8. (Suggested) Multi-Modal Perception [Lachs, 2017]
9. (Suggested) Decoding Brain Representations by Multimodal Learning of Neural Activity and Visual Features [Palazzo et al., 2020]
10. (Suggested) Similarity-Based Fusion of MEG and fMRI Reveals Spatio-Temporal Dynamics in Human Cortex During Visual Object Recognition [Cichy et al., 2016]

11. (Suggested) BRAINZOOM: High Resolution Reconstruction from Multi-modal Brain Signals [Fu et al., 2017]

We summarize several main takeaway messages from group discussions below:

1 Insights about the Brain

1.1 Brain Architecture

There has long been a hypothesis that the brain is a composition of multiple subsystems, each of which corresponds to some functionalities. In particular, O'Doherty et al. [2021] argues that the internal working of our brain is loosely analogous to that of the mixture of experts in machine learning. On the other hand, recent studies have shown that a functional organization of the human brain can be extended beyond the task it is specialized in. Norman and Thaler [2019] investigate how individuals with impaired vision can develop the ability to "see" through sound. We hypothesize that neurons in our brain might be general purpose, and our brain adopts weight-sharing and dynamic routing algorithms [Sabour et al., 2017]. However, we have fewer insights on whether the brain works in a centralized manner (i.e., there are one or more main nodes coordinating the subsystems) or in a decentralized manner (i.e., subsystems interact with each other directly without a coordinator).

1.2 Multimodality of the Brain

Nanay [2018] suggests that people often experience multimodal mental imagery triggered by sensory stimulation from a single modality. We further hypothesize that the brain perceives everything as multimodal and, similar to the SMIL system [Ma et al., 2021], it constantly reconstructs multimodal events with severely missing modalities from sensory stimulation. Another plausible theory is that the brain incorporates everything into one primary modality. One strong candidate for the primary modality is language. The language modality is more than just text - it can represent both low-level concepts grounded in other modalities and high-level ideas that are abstract. It can also be viewed as a discretized representation of thoughts, and we argue that the brain may use the language modality consciously or unconsciously as a "latent" modality.

1.3 Mental Imagery and Neural Signals

Multimodal mental imagery [Kosslyn et al., 2010] studies multimodal perception in the brain without actual perceptual stimulus. Mental imagery is normally involuntary and unconscious, involving neural pathways such as the occipital-temporal pathway, the occipital-parietal pathway, and the early visual cortex [Kosslyn et al., 2010].

Brain activity signals may also provide insights for multimodal learning. Nanay [2018], Kosslyn et al. [2010] study fMRI data and show that daily perception in the brain is multimodal. Anumanchipalli et al. [2019] explicitly leverages the kinematic and sound representations encoded in human cortical activity to synthesize audible speech. We argue that signals like fMRI can complement sensory modalities such as visual, acoustic, and textual input.

2 Brain-inspired AI Systems

2.1 Brain Versus AI systems

Although large pretrained models, for example GPT-3 [Brown et al., 2020], have been able to approach or even exceed human-level performance on many tasks, AI systems today are very different from our brain since they each excel at different sets of tasks. For example, our brain is good at generation [Teeple et al., 2009], modeling long-term interactions [Maheu et al., 2019], few-shot learning [Weaver, 2015], and creative tasks [Beaty et al., 2016], while AI systems in general may perform poorly on these tasks. There are other properties that distinguish the brain from existing AI systems. The human brain is large but very efficient, where different types of neurons and neurological pathways are activated for distinct tasks, for example, visual and ventral language pathways. These pathways or regions are dependent and spatially separated,

Brain	**ML models**
Is good at generation, long-term interaction, few-shot learning, and creative tasks	Needs large amounts of data and good training paradigms to perform well on these tasks
Has sparse neurons encoding discrete concepts	Most models use distributed representations only
Is robust to corrupted or noisy inputs	Needs techniques to improve representation if input is corrupted
Has neurons activated by mental imagery	Cannot be activated by mental imagery

Table 1: Attempts to understand the differences between the brain and existing ML models.

although there is some overlap [Eavani et al., 2015]. In general, neural activities can be sparse, and there is evidence that there is only one specific neuron for each visual concept [Gross, 2002]. Furthermore, the brain is robust to corrupted sensory input (e.g., occluded vision) and can be activated with mental imagery (i.e., without actual sensory input). We summarize these differences in Table 1.

2.2 Brain-Inspired AI Systems

The concept of multimodal imagery is similar to the idea of multimodal co-learning. Multimodal imagery generates perception in the brain for certain modalities, and the generated signals are used to help tasks in other modalities. We encourage readers to investigate whether co-learning actually happens in the brain, how it happens, what makes it possible to happen, and whether the current neural models are doing it the same way as the brain. We can then apply this understanding to design better multimodal systems facilitating co-learning.

Models should also be able to respond to imagery as well. Existing models depends on sensory input, like visual and acoustic input, and performs well on sensory oriented perception tasks, such as object recognition or VQA. But the actual brain performs differently: neurons in the brain can exhibit similar activation patterns from mental imagery as if the actual input were there, even if the actual input is not present. Future studies should investigate the role of imagery in forming mental representation states and how this mechanism can help us design better machine learning systems.

2.3 Comparing Models with the Brain

More research needs to be done to assess how closely activations in the human brain and AI systems align. For example, fMRI or Neuralink mesh could be used to provide a fine-grained analysis of how neurons in our brain are activated and updated, which can then be compared with AI systems performing similar tasks.

Another way of evaluation is to compare task-specific neurons in the brain vs neurons in AI models. Deep Sparse Coding [Kim et al., 2018] takes inspiration from the brain to make neurons more task-specific, and we could compare the neural activation of neurons in the Deep Sparse Coding model and the actual neurons in the brain.

3 Future Work

Studying the human brain provides opportunities to improve current AI systems. In addition to the research ideas discussed above, we list a few more directions for future work:

- It will be helpful to first create a taxonomy of the differences between AI models and the brain. This will give insights into how current AI systems can be improved.
- To help an AI system learn better from human action, we can enable AI systems to observe signals from the human brain by using brain signals as another modality to assist learning.

References

Gopala K Anumanchipalli, Josh Chartier, and Edward F Chang. Speech synthesis from neural decoding of spoken sentences. *Nature*, 568(7753):493–498, 2019.

Roger E Beaty, Mathias Benedek, Paul J Silvia, and Daniel L Schacter. Creative cognition and brain network dynamics. *Trends in cognitive sciences*, 20(2):87–95, 2016.

Manuel Blum and Lenore Blum. A theoretical computer science perspective on consciousness. *Journal of Artificial Intelligence and Consciousness*, 8(01):1–42, 2021.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

Gemma A Calvert. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral cortex*, 11(12):1110–1123, 2001.

Radoslaw Martin Cichy, Dimitrios Pantazis, and Aude Oliva. Similarity-based fusion of meg and fmri reveals spatio-temporal dynamics in human cortex during visual object recognition. *Cerebral Cortex*, 26(8): 3563–3579, 2016.

Harini Eavani, Theodore D Satterthwaite, Roman Filipovych, Raquel E Gur, Ruben C Gur, and Christos Davatzikos. Identifying sparse connectivity patterns in the brain using resting-state fmri. *Neuroimage*, 105: 286–299, 2015.

Xiao Fu, Kejun Huang, Otilia Stretcu, Hyun Ah Song, Evangelos Papalexakis, Partha Talukdar, Tom Mitchell, Nicholas Sidiropoulo, Christos Faloutsos, and Barnabas Poczos. Brainzoom: High resolution reconstruction from multi-modal brain signals. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, pages 216–227. SIAM, 2017.

Charles G Gross. Genealogy of the "grandmother cell". *The Neuroscientist*, 8(5):512–518, 2002.

Edward Kim, Darryl Hannan, and Garrett Kenyon. Deep sparse coding for invariant multimodal halle berry neurons. In *Proceedings of the ieee conference on computer vision and pattern recognition*, pages 1111–1120, 2018.

Stephen M Kosslyn, Giorgio Ganis, and William L Thompson. Multimodal images in the brain. *The neurophysiological foundations of mental and motor imagery*, pages 3–16, 2010.

Lorin Lachs. Multi-modal perception. *Noba textbook series: Psychology. Champaign: DEF Publishers*, 2017.

Mengmeng Ma, Jian Ren, Long Zhao, Sergey Tulyakov, Cathy Wu, and Xi Peng. Smil: Multimodal learning with severely missing modality. *arXiv preprint arXiv:2103.05677*, 2021.

Maxime Maheu, Stanislas Dehaene, and Florent Meyniel. Brain signatures of a multiscale process of sequence learning in humans. *elife*, 8:e41541, 2019.

Bence Nanay. Multimodal mental imagery. *Cortex*, 105:125–134, 2018.

Liam J Norman and Lore Thaler. Retinotopic-like maps of spatial sound in primary 'visual'cortex of blind human echolocators. *Proceedings of the Royal Society B*, 286(1912):20191910, 2019.

Magalie Ochs, Roxane Bertrand, Aurélie Goujon, Deirdre Bolger, Anne-Sophie Dubarry, and Philippe Blache. The brain-ihm dataset: a new resource for studying the brain basis of human-human and human-machine conversations. In *Language Resources and Evaluation Conference (LREC)*, 2020.

John P O'Doherty, Sang Wan Lee, Reza Tadayonnejad, Jeff Cockburn, Kyo Iigaya, and Caroline J Charpentier. Why and how the brain weights contributions from a mixture of experts. *Neuroscience & Biobehavioral Reviews*, 123:14–23, 2021.

Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, Joseph Schmidt, and Mubarak Shah. Decoding brain representations by multimodal learning of neural activity and visual features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11):3833–3849, 2020.

Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. *Advances in neural information processing systems*, 30, 2017.

Dan Schwartz, Mariya Toneva, and Leila Wehbe. Inducing brain-relevant bias in natural language processing models. *Advances in neural information processing systems*, 32, 2019.

Ryan C Teeple, Jason P Caplan, and Theodore A Stern. Visual hallucinations: differential diagnosis and treatment. *Primary Care Companion to the Journal of Clinical Psychiatry*, 11(1):26, 2009.

Janelle Weaver. How one-shot learning unfolds in the brain. *PLoS biology*, 13(4):e1002138, 2015.