Neurodegenerative Modelling and Computer Vision: The Role of Attention Mechanisms

Final Paper

PS467 - Seminar in Cognitive Neuroscience

Wilfrid Laurier University

April 3, 2023

Prepared for: Dr. Philip Servos

Prepared by: Neekesh Panchal

Neurodegenerative Modelling and Computer Vision: The Role of Attention Mechanisms

In the developing field of computer vision (CV), there has been a recent steep incline of researchers working on improving novel approaches (Convolutional Neural Networks, LSTMs, Autoencoders, RNNs) through reverse engineering neural attention processes. An example of this emerged in 2017 when a group of researchers successfully released a purely attention mechanism-based model called a Transformer (Vaswani et al., 2017). This landmark model introduced attention-based learning - as observed in cognitive science - towards implementing an informed (and highly complex) predictive architecture. Many attribute the attention-head mechanisms within transformer architectures to advances in neuroscience, however, the two innovative approaches emerged in parallel. This paper builds upon and extracts critical findings from previous literature, emphasizing attentional mechanisms in the primary visual area (V1) and its many downstreams (V2, V3, V4). It begins by reviewing early information of attention-based mechanisms in the brain and where this research converged with novel approaches in CV. This paper then considers updating current CV architectures for neurodegenerative modelling through recent advances in neural attention mechanics. The aim is to discuss in-depth why attention mechanisms are important for visual information processing, the contrasting approaches in CV and cognitive neuroscience, and how researchers can leverage recent strides in neural attention mechanisms towards improving CV models for neurodegenerative disease modelling and analysis.

Section A: Neural Attention Mechanisms in Visual Information Processing

The intricate role of attention in visual information processing, particularly within the occipital lobe, forms a cornerstone in our understanding of human perception. The field of CV requires us to reverse-engineer these mechanics of the occipital lobe towards developing computer software which provide similar functionality. Attention mechanisms are not merely passive conduits of visual stimuli, but are active processes which filter, enhance, and interpret sensory information. These attention mechanisms communicate with neurons to elicit changes in their reactivity. They are heavily nuanced as the active processes can vary due to several factors, including quantal synaptic dynamics, multimodal integrative noise, and even parallel implicit processes such as perceptual learning. Within this section, I will review earlier literature which firstly emphasizes the neural substrate of attention-based mechanisms within visual integration areas, as well as explore the potential nuances of how these mechanisms affect the properties of a neuron.

First define whether attention-based mechanisms alter occipital neuronal responses, then dive into the specific processes at a lower-level. Crist et al. (2001) elucidate the dynamics of attention in the primary visual cortex (V1), a critical region within the occipital lobe. Their findings reveal that attention modulates neural activity in area V1, not by altering the receptive field's size but by enhancing the response of neurons to stimuli within their receptive fields. This enhancement leads to a more precise and focused processing of visual information. The study underscores that attention does not act as a mere spotlight illuminating areas of interest but rather as a sophisticated mechanism that sharpens and refines what is already being visually perceived. This "enhancement" opposes an earlier notion that the receptive field size is the neuronal alteration which occurs after receiving an attention-based signal. As researchers statistically rejected this hypothesis, the changes were more nuanced than only considering the impact on receptive size.

Based on the work by Crist et al. in 2001, attention-based mechanisms modulate neural activity in area V1 through so-called enhancive and diminutive processes. An important assumption made here is that attention has an entirely multimodal effect. Researchers have experimented and tested this assumption, yielding results which support the hypothesis of multimodal effects of attention. Sprague & Serences (2013) employ novel computational techniques such as autoencoders and ROI points to dissect patterns in modality response to attention-guided tasks. Through analyzing the fMRI data generated while participants were engaged in a visual task, researchers reconstructed a spatial representation of areas V1, extrastriate, parietal, and frontal cortical activation patterns. They found that the spatial representations of these visual stimuli were incredibly accurate in area V1 but became gradually more diffuse in the extrastriate, parietal, and frontal cortices (Sprague & Serences., 2013). An effect of attention-based response is observed at each region, such that the neural responses observed were directly related to the attention-based stimulation. The progressive diffusive nature insinuates that attentional mechanisms vary across the cortical areas of interest. The extrastriate and regions other than area V1 typically process higher-dimensional information, implying that the spatial reconstruction was not as accurate due to more complex integration.

This section concludes by analyzing two main understandings within the relevant literature. The first understanding is that attention-based mechanisms affect neural response dynamics such that enhancive and diminutive states guide perpetual alterations in neuron structure. The second understanding being that attention is integrated across modalities, and that area V1 is highly accurate at decoding the spatial layout of visual information. These two points equip us with the necessary background information to begin discussing why and how attention mechanisms in the human occipital lobe have been computationally leveraged to assist those with neurodegenerative complications.

Section B: Parallels between Computer Vision & Cognitive Science Attention Mechanics

In computer vision, attention mechanisms direct the focus of neural networks towards relevant segments of visual data (Guo et al., 2022). Through mimicking aspects of human selective attention, they enable models to prioritize information-rich areas in images or video frames, thus improving tasks like object recognition and spatial layout. Attention mechanisms in cognitive science are incredibly complex and nuanced due to hidden variables which influence neuronal changes. Due to this, contemporary attention-based mechanics in computer vision replicate seemingly simple yet computationally intensive occipital system properties. Leveraging the transformer architecture mentioned earlier and accompanying it with visual data processing capabilities resulted in a novel variation called a Vision Transformer (ViT). A ViT is much more complex and uses temporal attention mechanics, which can produce remarkably coherent and accurate analytics from images and videos. This new ViT architecture now integrates attention-based functionality present not only within our occipital but from the temporal lobe as well. This section covers exploration on how a ViT works, how the processes of a ViT are analogous to attention mechanics in the brain, and the current limitations of analogies between the domains.

A ViT introduces a novel approach by adapting the Transformer architecture, traditionally used in natural language processing (NLP), to visual data. In a ViT, an image is partitioned into fixed-size patches, akin to words in a sentence (Courant et al., 2023). Each patch is then flattened and transformed into a vector, representing the token for the transformer. This tokenization allows the model to process visual information like text data in NLP (2023). The core of a ViT is its attention mechanism, which employs learnable weights to compute attention scores between all pairs of tokens. The notorious scaled dot-product attention from the Transformer architecture is utilized here. It computes the dot product between each token (query) and all other tokens (keys), scales the value, and then applies softmax (activation function) to determine the final attention weights. These weights signify the relative importance of each token to the others. For output generation, the ViT calculates a weighted sum of the tokens' vectors (values), guided by the attention weights. This process allows the network to focus selectively on more informative parts of the image.

The methods outlined prior in which ViTs deconstruct, and process visual input are analogous to high-level attention mechanisms observed in humans. In human vision, neurons in area V1 selectively process visual stimuli, focusing on critical regions and filtering out extraneous information. This biological mechanism closely mirrors the Vision Transformers' (ViT) functionality, where varying attention weights are assigned to different image patches to determine their relative importance (Khatri & Kwon, 2023). The dot-product attention of ViTs is analogous to how neurons in our brain might strengthen synaptic connections for significant stimuli. In addition, layer normalization and feed-forward networks in ViTs reflect the neuronal process of normalizing and integrating signals. ViT functions clearly parallel what was discussed earlier, the enhancive and diminutive dynamics within neurons of the occipital lobe and how these neurons adjust themselves based on information passed to them from these attention transients (Crist et al., 2001). In essence, just as our brain's neurons selectively exacerbate or diminish signals to direct attention, ViTs algorithmically accentuate specific aspects of visual data, effectively emulating natural attention mechanisms through high-dimensional vector manipulation.

CV approaches functionally perform elementary tasks in contrast to more high-dimensional problems our human visual processing system is capable of handling. Researchers are currently limited in replicating our visual cortices' attention capabilities. To reverse-engineer these processes, one must first understand how to define their properties. Contemporary research in cortical attention mechanisms continues to evolve and, in parallel, shape the future path for computer vision. For example, a recent study from 2023 investigated the effects of suppressing feedback from the V4 area to V1 in monkeys performing attention tasks. The suppression led to a significant decrease in attentional modulation, impacting response gain and the accuracy of neural responses, especially in the superficial layers of V1 (Debes & Dragoi, 2023). These findings indicate a novel effect of attention mechanics on neural responses, an intriguing dynamic and attentional process not considered anywhere within the ViT architecture. Considering this effect of suppressing layer communication, implementing similar parameters to fine-tune a ViT and testing this functionality could result in a more descriptive model. Over time, as the field matures, researchers hope to see these newly uncovered attentional mechanics being reverse engineered into CV algorithms, assisting ViTs in becoming more robust.

Section C: Neurodegenerative Models and Attention Mechanisms

Neurodegenerative diseases like Alzheimer's, Parkinson's, and Huntington's are characterized by their complex, progressive nature and often subtle initial symptoms. This makes early detection incredibly challenging yet crucial for effective treatment. CV models, particularly those enhanced with attention mechanisms inspired by human visual processing, have shown the potential to reliably identify and analyze key disease markers in neuromedical imaging, such as MRI and PET scans. These models offer novel approaches towards early disease detection and progression monitoring by extracting and interpreting complex patterns imperceptible to the human eye. Within this section, it begins with discussing contemporary use cases of CV algorithms, then delves into how novel understandings of neural attention mechanisms can influence the future research direction of ViTs.

Alzheimer's and Parkinson's diseases have been central in case studies involving CV models. For Alzheimer's, CV models analyze brain scans to detect features like amyloid plaques and neurofibrillary tangles (Johnson et al., 2012). Parkinson's disease studies focus on changes in the substantia nigra and related structures by applying deformation fields to neuroimaging data (Deng et al., 2022). These approaches in CV employ algorithms like Convolutional Neural Networks (CNNs), which are increasingly being outclassed by Vision ViTs. ViTs, with their attention mechanisms, are incredibly adept at focusing on relevant portions of the medical images, discerning subtle changes in brain volume, white matter integrity, and other critical biomarkers indicative of early stages of neurodegenerative diseases.

Despite researchers' achievements in applying Vision Transformers (ViTs) to neuroimaging analysis, achieving high accuracy and reliability remains a steep challenge. Insights from neuroscience studies on attention mechanisms provide valuable guidance for refining these models and extending the range of their capabilities. For instance, the research conducted by Braun et al. (2002) revealed that attention actively and dynamically influences the way the primary somatosensory cortex represents the hand. This modulation is not static but varies depending on the specific task being performed, suggesting that attention allocation reshapes neural responses. In the context of ViTs, this principle can be applied by developing models that dynamically adjust their focus based on the semantic processing of the task, potentially enhancing multimodal models toward accurately detecting and interpreting disease markers in neurodegenerative diseases.

Farrer and Passingham (2002) illustrated that tasks requiring explicit recognition and response association activate the ventral visual stream. This finding implies that higher cognitive functions like explicit recognition and decision-making utilize specific neural pathways. In ViT development, this suggests a move towards models that do more than solely detect features in medical images. Instead, these models could associate these features with specific neurodegenerative markers, enhancing their interpretative capabilities. For example, in a ViT model designed for Alzheimer's disease, based on Farrer & Passingham's findings, one could implement a mechanism where the model identifies amyloid plaques in PET scans and correlates their spatial distribution with disease progression stages. This approach, akin to the ventral stream's role in associating visual stimuli with responses, would allow for a more nuanced analysis of the imaging data. Similarly, applying the insights from Braun et al., ViTs could be adapted to dynamically focus on different brain regions based on the specific neurodegenerative condition being analyzed. For instance, in Parkinson's disease, the model could adjust its focus to the substantia nigra and related structures, dynamically changing its attention based on the task – akin to the modulations seen in the primary somatosensory cortex as reported by Braun et al.

In conclusion, this paper explored the intersection between computer vision and cognitive neuroscience, particularly focusing on the role of attention mechanisms in neurodegenerative disease modelling. The sections delved into insights from neural attention studies that can be applied to enhance computer vision models, such as Vision Transformers, for a more accurate and nuanced analysis of neurodegenerative conditions. By integrating these interdisciplinary approaches, researchers hope to significantly advance diagnostic and monitoring capabilities, allowing for earlier detection and potentially more effective treatment strategies for progressive diseases like Alzheimer's and Parkinson's. As the field of computer vision matures, there will be a surge of novel software implementations which synthesize the capabilities of human attention mechanics into versatile and reliable computer vision algorithms.

References

Arno, P., De Volder, A., Vanlierde, A., Wanet-Defalque, M., Streel, E., Robert, A., Sanabria-Bohórquez, S., & Veraart, C. (2001). Occipital activation by pattern recognition in the early blind using auditory substitution for vision. NeuroImage. https://pubmed.ncbi.nlm.nih.gov/11305892/

Braun, C., Haug, M., Wiech, K., Birbaumer, N., Elbert, T., & Roberts, L. (2002). Functional organization of primary somatosensory cortex depends on the focus of attention. NeuroImage. https://pubmed.ncbi.nlm.nih.gov/12414284/

Courant, R., Edberg, M., Dufour, N., Kalogeiton, V., & Colliot, O. (2023). Transformers and visual transformers. National Center for Biotechnology Information. https://pubmed.ncbi.nlm.nih.gov/37988536/

Crist, R. E., Li, W., & Gilbert, C. D. (2001). Learning to see: Experience and attention in primary visual cortex. Nature News. https://www.nature.com/articles/nn0501_519

Debes, S., & Dragoi, V. (2023). Suppressing feedback signals to visual cortex abolishes attentional modulation. Science (New York, N.Y.). https://pubmed.ncbi.nlm.nih.gov/36730414/

Deng JH, Zhang HW, Liu XL, Deng HZ, Lin F. (2022). Morphological changes in Parkinson's disease based on magnetic resonance imaging: A mini-review of subcortical structures segmentation and shape analysis. *World J Psychiatry*, *12*(12), 1356-1366. https://doi.org/10.5498/wjp.v12.i12.1356

Farrer, C., Passingham, R., & Frith, C. (2002). A role for the ventral visual stream in reporting movements. NeuroImage. https://pubmed.ncbi.nlm.nih.gov/11848701/

Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., Zhang, S.-H., Martin, R. R., Cheng, M.-M., & Hu, S.-M. (2022 March). Attention mechanisms in computer vision: A survey - computational visual media. SpringerLink. https://link.springer.com/article/10.1007/s41095-022-0271-y#Ack1

Johnson, K. A., Fox, N. C., Sperling, R. A., & Klunk, W. E. (2012). Brain imaging in Alzheimer disease. Cold Spring Harb *Perspect Med*, *2*(4). https://doi.org/10.1101/cshperspect.a006213

Khatri, U., & Kwon, G. R. (2023). Explainable Vision Transformer with Self-Supervised Learning to Predict Alzheimer's Disease Progression Using 18F-FDG PET. *Bioengineering* (Basel, Switzerland), *10*(10), 1225. https://doi.org/10.3390/bioengineering10101225

Sprague, T. C., & Serences, J. T. (2013, November 10). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nature News*. https://www.nature.com/articles/nn.3574

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017, January). Attention is all you need. *Advances in Neural Information Processing Systems*. https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845a a-Abstract.html