Visual Attention in Artificial and Biological Neural Networks

Grace Lindsay, PhD

Research Fellow, Gatsby Computational Neuroscience Unit/ Sainsbury Wellcome Centre
Outline

• Some concepts of attention in psychology/neuroscience and how they relate to machine learning
• Using convolutional neural networks to understand feature-based attention in the brain
• Making CNNs more biologically realistic
Attention in Psychology, Neuroscience, and Machine Learning

Grace W. Lindsay

Gatsby Computational Neuroscience Unit, Sainsbury Wellcome Centre, University College London, London, United Kingdom

1. Attention in neuroscience and psychology

2. Attention in machine learning, with similarities to biological attention indicated

3. Ideas for future interaction between artificial and biological attention
Attention: the ability to flexibly control limited computational resources
Attention as overall arousal

Yerkes-Dodson Curve

Performance

Arousal
Attention as overall arousal

Yerkes-Dodson Curve

Neocortex
- Sensory processing, executive planning, motor control

Amygdala
- Fear & anxiety

Hippocampus
- Memory & navigation

Basal Forebrain
- Sends cholinergic projections

Thalamus
- Sensory gating

Hypothalamicus
- Homeostasis

Locus Coeruleus
- Releases norepinephrine

Cerebellum
- Motor coordination & sensory prediction

Brain Stem
- Autonomic control

Spinal Cord
- Sensory input & motor output
Saccades

- Small eye movements made several times per second

**Fig. 1.** Huey’s [10] lever device to record horizontal eye movements. *a* Eye movements made during reading were recorded with this technique; from Huey [11]. *b* The tracing on the smoked drum was photographed and then engraved; from Wade et al. [1].

Eggert, 2007
Saccades

- “Overt spatial attention”: limited computational resource, controlled flexibly

Eye trajectories measured by Yarbus by viewers carrying out different tasks. (a) No specific task. (b) Estimate the wealth of the family. (c) Give the ages of the people in the painting. (d) Summarize what the family had been doing before the arrival of the “unexpected visitor”. (e) Remember the clothes worn by the people. (f) Remember the position of the people and objects in the room. (g) Estimate how long the “unexpected visitor” had been away from the family. Image adapted from Yarbus (1967)
Covert Spatial Attention

- Overt spatial attention remains fixed
- Valid cueing enhances performance
Covert Spatial Attention

Neural activity is modulated based on preferred spatial location (red lines = attending into a cell’s receptive field, blue line = outside)

Buffalo et al., 2010
“Hard” Spatial Attention
“Hard” Spatial Attention

- Train a recurrent CNN with reinforcement learning to select image regions for further processing.

Mnih, et al., 2014
“Soft” spatial attention

- Iterative reweighting of hidden layer activity, can be trained with backprop

“Show, Attend and Tell: Neural Image Caption Generation with Visual Attention”, Xu et al., 2015
Covert feature-based attention

Lupyan and Ward, 2013
How biological attention mechanisms improve task performance in a large-scale visual system model

Grace W Lindsay¹,²*, Kenneth D Miller¹,²,³,⁴

¹Center for Theoretical Neuroscience, College of Physicians and Surgeons, Columbia University, New York, United States; ²Mortimer B. Zuckerman Mind Brain Behaviour Institute, Columbia University, New York, United States; ³Swartz Program in Theoretical Neuroscience, Kavli Institute for Brain Science, New York, United States; ⁴Department of Neuroscience, Columbia University, New York, United States
How can we explore the connection between the neural changes that accompany attention and performance changes?
Connecting Neural Changes with Performance Changes Using Convolutional Neural Networks

- Architecture is inspired by the visual system
Connecting Neural Changes with Performance Changes Using Convolutional Neural Networks

- Architecture is inspired by the visual system

Herzog and Clarke, 2014
Connecting Neural Changes with Performance Changes Using Convolutional Neural Networks

- Representations are similar too.

VGG-16

Fully Connected (1000)
Fully Connected (4096)
Fully Connected (4096)
Max-Pooling
Convolution (512)
Convolution (512)
Convolution (512)
Max-Pooling
Convolution (512)
Convolution (512)
Convolution (512)
Max-Pooling
Convolution (256)
Convolution (256)
Convolution (256)
Max-Pooling
Convolution (128)
Convolution (128)
Max-Pooling
Convolution (64)
Convolution (64)
Image (224x224x3)
Covert feature-based attention enhances performance in challenging detection tasks.

VGG-16

Fully Connected (1000)
Fully Connected (4096)
Fully Connected (4096)

Max-Pooling

13 Convolution (512)
12 Convolution (512)
11 Convolution (512)
10 Convolution (512)
9 Convolution (512)
8 Convolution (512)
7 Convolution (256)
6 Convolution (256)
5 Convolution (256)
4 Convolution (128)
3 Convolution (128)
2 Convolution (64)
1 Convolution (64)

Image (224x224x3)

Lupyan and Ward, 2013
VGG-16

Fully Connected (1000)
Fully Connected (4096)
Fully Connected (4096)

Max-Pooling

Convolution (512)
Convolution (512)
Convolution (512)

Max-Pooling

Convolution (512)
Convolution (512)
Convolution (512)

Max-Pooling

Convolution (256)
Convolution (256)
Convolution (256)

Max-Pooling

Convolution (128)
Convolution (128)

Max-Pooling

Convolution (64)
Convolution (64)

Image (224x224x3)

Binary Classifier: “Clock”
Fully Connected (4096)
Fully Connected (4096)

Binary Classifier: “Greenhouse”
Fully Connected (4096)
Fully Connected (4096)

Test Images

MERGED

ARRAY
VGG-16

Binary Classifier: “Clock”
- Fully Connected (4096)
- Fully Connected (4096)

Binary Classifier: “Greenhouse”
- Fully Connected (4096)
- Fully Connected (4096)

Box plots showing binary classification performance for Standard, Merged, and Array categories.
Neural Correlates of Attention
Neural Correlates of Attention

• “Feature Similarity Gain Model”: Increased firing rate for preferred targets, decreased for anti-preferred. Effects are multiplicative.

Maunsell and Treue, 2006

Treue, 2001
Neural Correlates of Attention

• “Feature Similarity Gain Model”: Increased firing rate for preferred targets, decreased for anti-preferred. Effects are multiplicative.

• Feature-based attention is spatially global
Neural Correlates of Attention

• “Feature Similarity Gain Model”: Increased firing rate for preferred targets, decreased for anti-preferred. Effects are multiplicative.

• Feature-based attention is spatially global

• Effects tend to be stronger at later areas
Modeling Attention

- Replicating the “feature similarity gain model”

1.) Make normalized category tuning curves for each feature map:

2.) When attention is applied to a category, the activity is scaled according to the tuning for that category:

\[ x_{lk}^{ij} = (1 + \beta f_{lk}^c)[I_{lk}]_+ \]
Do neural manipulations like those observed biologically enhance performance?
Do neural manipulations like those observed biologically enhance performance?
How *should* activity be modulated?

Tuning values reflect how a feature map’s activity relates to the network’s input (here, images of different categories).
How should activity be modulated?

Tuning values reflect how a feature map’s activity relates to the network’s input (here, images of different categories).

Gradient values reflect how a feature map’s activity relates to the network’s output (here, different category labels).
How *should* activity be modulated?

Tuning values reflect how a feature map’s activity relates to the network’s input (here, images of different categories).

Gradient values reflect how a feature map’s activity relates to the network’s output (here, different category labels).
Object Category Detection Task:

How should activity be modulated?
How should activity be modulated?

- Applying attention according to gradient values leads to better performance than tuning values at early layers.
How *should* activity be modulated?

- It's unclear whether brain uses something like gradient values.
Soft feature attention

A more biologically detailed approach
A more biologically detailed approach

A convolutional neural network was merged with an existing model of visual cortex to include:

- Separate excitatory and inhibitory cells
- Supralinear activation function
- Distance-dependent recurrent connectivity
- Realistic neural temporal dynamics
A more biologically detailed approach
Summary

• The feature similarity gain model of attention is effective at enhancing performance in CNNs

• Calculating how attention should modulate activity ('gradient values') leads to better performance

• Whether the brain is using tuning or gradient values is unclear

• A more biologically-detailed model can replicate these findings
Thanks!