Some notes on Deep Hierarchies in Human/Biological Vision

CEng 783 – Deep Learning
Fall 2017

Emre Akbaş
• Project final presentations are next week.
  – Groups expected to present were announced.
• Project report/paper is due Jan 14.
The figure illustrates the distribution of rod and cone densities across different areas of the retina. The x-axis represents the perimeter angle, ranging from 70° to 80°, with the fovea at 0°. The y-axis shows the density of rods or cones per mm², ranging from 20k to 180k.

- **Fovea**: High density of cones with a dense packing of cone photoreceptors, indicating a high visual acuity region.
- **Blind spot**: Transition zone with lower photoreceptor density, marking the point where the optic nerve exits the eye.
- **Temporal area** and **Nasal area**: Increasing density of rods towards the peripheral areas, with cones decreasing.

Inset images depict the structure of the fovea and periphery, highlighting the difference in photoreceptor distribution.
The output of the eye = The output of retinal ganglion cells
Retinal Ganglion Cells

Two main types:
ON-center-OFF-surround and OFF-center-ON-surround

Cell 1

Cell 2

Stephen Kuffler – 1957

[From Ch. 3 of Hubel (1995)]
After the RGCs?

Between the retinal ganglion cells (RGCs) and the visual cortex:

- Lateral Geniculate Nucleus: simply relays RGC output to the primary visual cortex.
The primary visual cortex

Also known as “the area V1” or “the striate cortex.”

David Hubel and Torsten Wiesel won the Nobel prize for discovering the functional organization and basic physiology of neurons in V1.
V1 physiology: orientation selectivity

[Slide by David Heeger]
V1 physiology:

direction selectivity

[Slide by David Heeger]
V1 simple cells

Simple cell

[Slide by David Heeger]
V1 complex cell

Complex cell

[Slide by David Heeger]
How is the wiring?

Simple cell sums LGN inputs
The concept of “receptive field”

The receptive field of an individual neuron is the particular region of the visual field where a stimulus will modify the firing of that neuron.

[From Ch. 3 of Hubel (1995)]
The resolution of the human visual system is not uniform across the visual scene.
Physiological measurements of receptive field size in macaque. (a) Receptive field size (diameter) as a function of the distance between the receptive field center and the fovea (eccentricity) for visual areas V1, V2 and V4. (b) Cartoon depiction of receptive fields with sizes based on physiological measurements.

[Freeman & Simoncelli, 2011]
The Freeman-Simoncelli Model

Specifies the locations, sizes and functions of cells in V1, V2
• An original image (b) is passed through the model (a).
• Synthetic images can be sampled/generated such that their model responses are matched to that of (b).
• Show two random synthetic image X and Y, and then as the third show either X or Y, and ask: is the third X or Y?
The fact that we do not see our surroundings in uniform, high resolution manifests itself in visual search.
Visual search
Visual search

How does computer vision solve this problem?
Visual search
Visual search

Yellow circles are called fixation points.

Saccade: rapid movement of the eyes between fixation points.
Human visual system: foveal vs peripheral vision

Fovea is responsible for tasks that need fine spatial detail such as reading and driving.
Using the Freeman-Simoncelli Model for computer vision object detection

- Sliding window vs foveated search
Operation of the foveated model in one cycle:

1) Fixate at f
2) Extract V1 complex cell features
3) Compute (and record) posterior probabilities at each location
4) Find the location with max prob and set it as the next fixation location
5) Go to step 1 if number of fixations did not exceed a given value.
Results

[ Akbas & Ecsktein, 2017 ]

\[ \text{~30x speedup} \]
Microsaccades
Beyond V1?

Highly recommended reading: *Deep hierarchies in the primate visual cortex: What can we learn for computer vision?* [Kruger et al., 2012]
- CNN representations explain ventral stream fMRI.
- This mapping follows the known hierarchical organisation.
- A full brain predictive model synthesizes brain maps for other visual experiments.
- Only deep models can reproduce observed BOLD activity.
- **Top:** ventral visual pathway, **bottom:** a CNN
- **Results:**
  - Top hidden layers of CNNs predict IT neuronal responses.
  - Intermediate and lower layers predict V4 and V1 responses.

[Yamins & DiCarlo, 2016]
Compared temporal (magnetoencephalography) and spatial (functional MRI) visual brain representations with representations in an artificial deep neural network (DNN).

They showed that the DNN captured the stages of human visual processing in both time and space from early visual areas towards the dorsal and ventral streams.
A fun example to draw contrast between human vision and computer vision in object detection:

Find the two toothbrushes in the following image.
References


References

