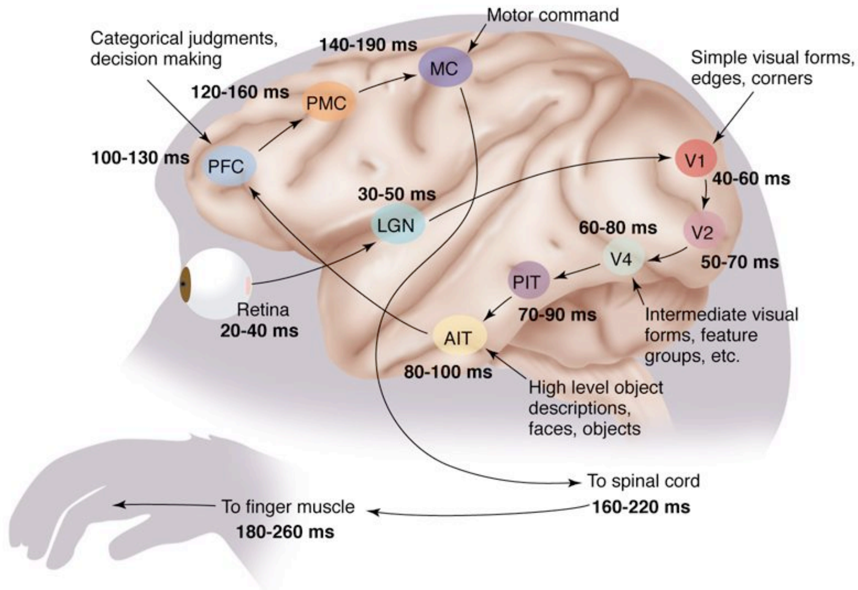
EYESHOTS
Kick-off Meeting

Attention and 3D-Object recognition

PD Dr. Fred Hamker

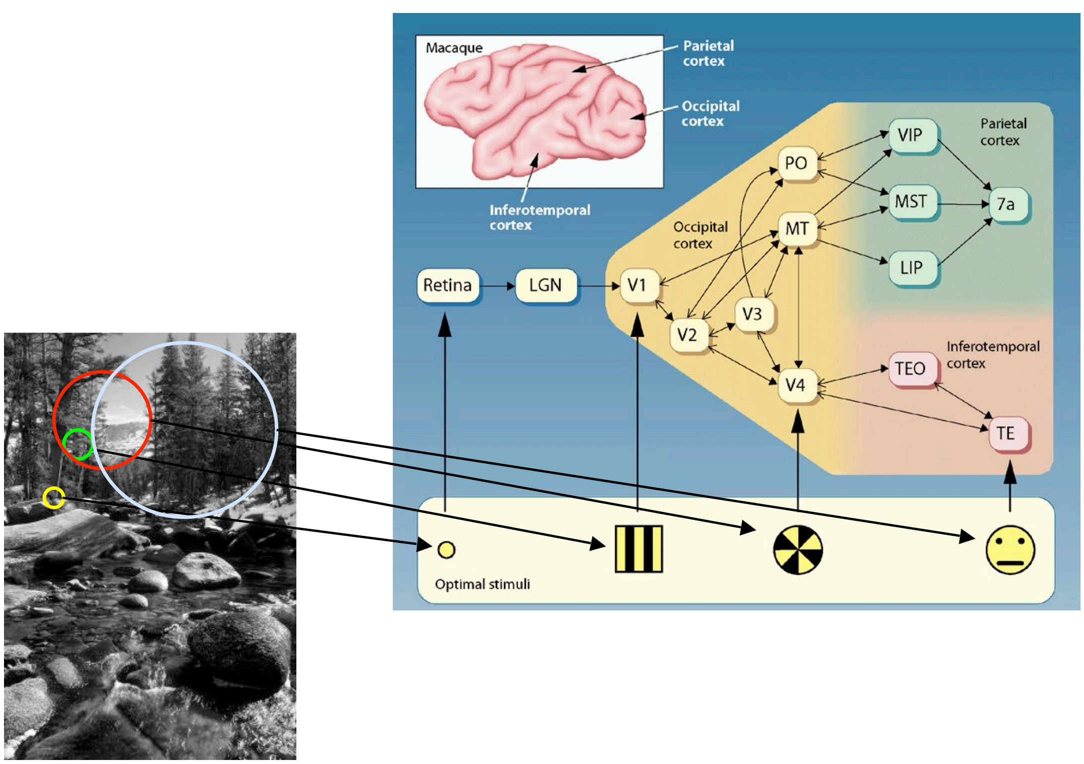
Westf. Wilhelms-University Münster, Germany
Department of Psychology
Otto-Creutzfeldt Center for Cognitive and Behavioral Neuroscience

Object recognition in primates



Thorpe, SJ & Fabre-Thorpe, M (2001)
Seeking categories in the brain. Science, 291:260-263.

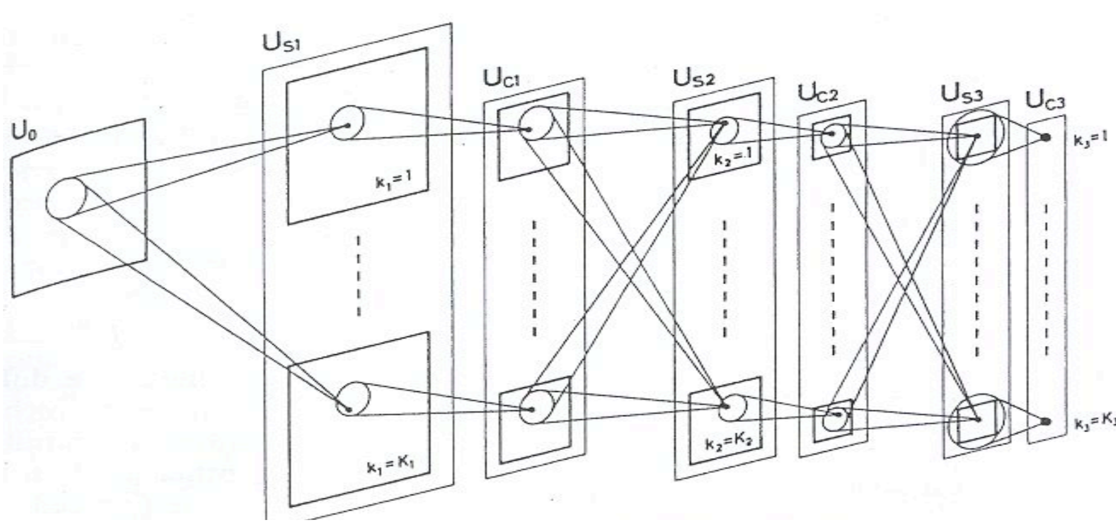
The ventral and dorsal pathway



Models of object recognition

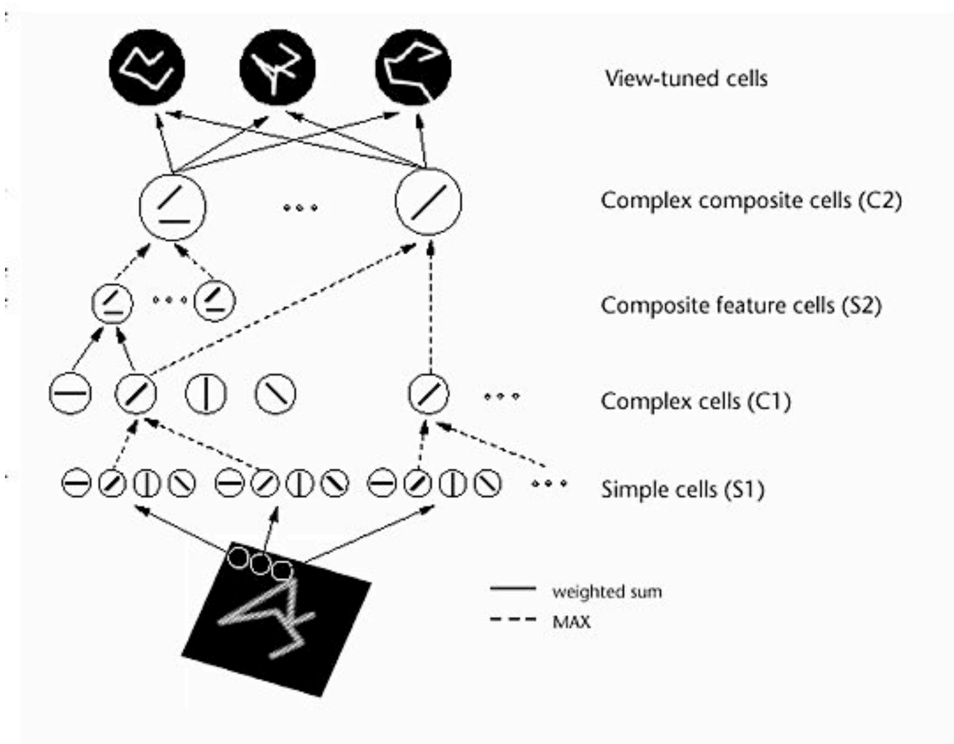
- Fukushima, K (1980) Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biol. Cybern. 36:193-202.
- Riesenhuber, M & Poggio, T (1999) Hierarchical models of object recognition in cortex, Nat. Neurosci. 2:1019-1025.

Position invariant recognition in the Neocognitron (Fukushima 1980)

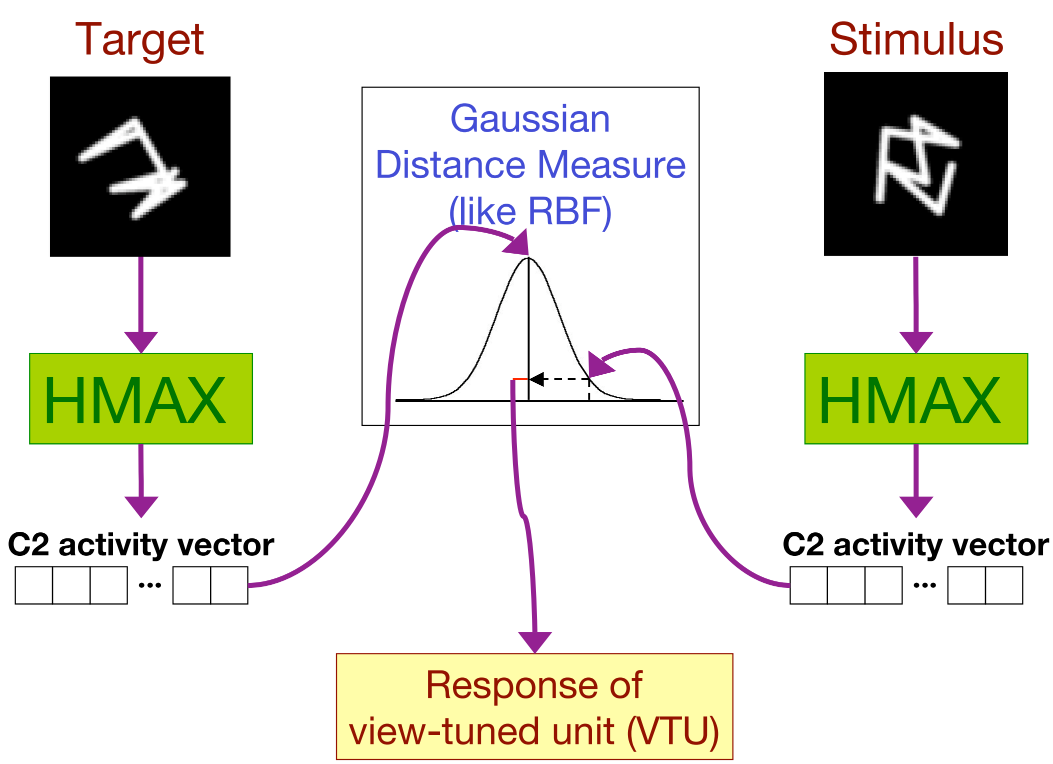


Several processing layers, comprising simple (S) and complex (C) cells.
S-cells in one layer respond to conjunctions of C-cells in previous layer.
C-cells in one layer are excited by small neighborhoods of S-cells.

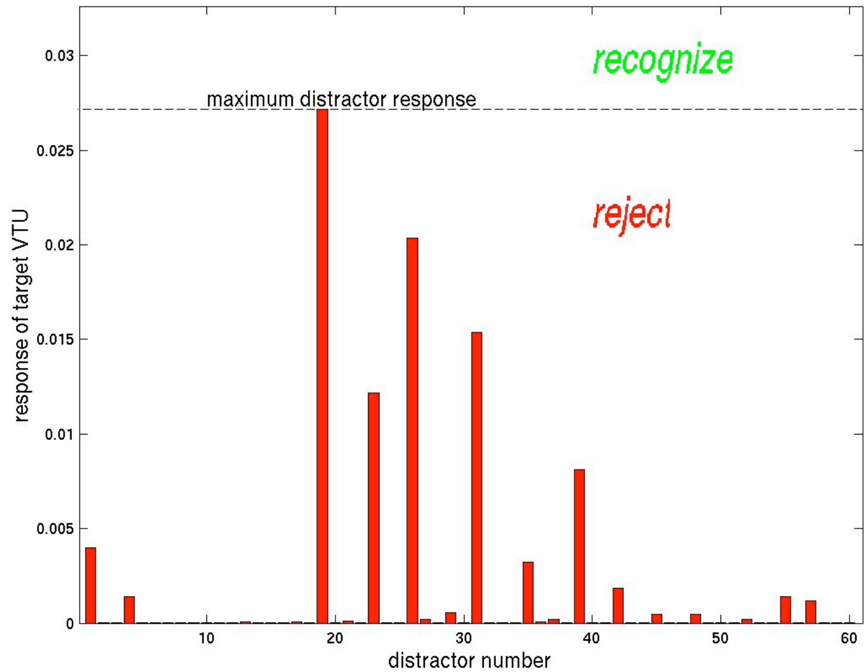
HMAX (Riesenhuber and Poggio, 1999)



Details of recognition in HMAX

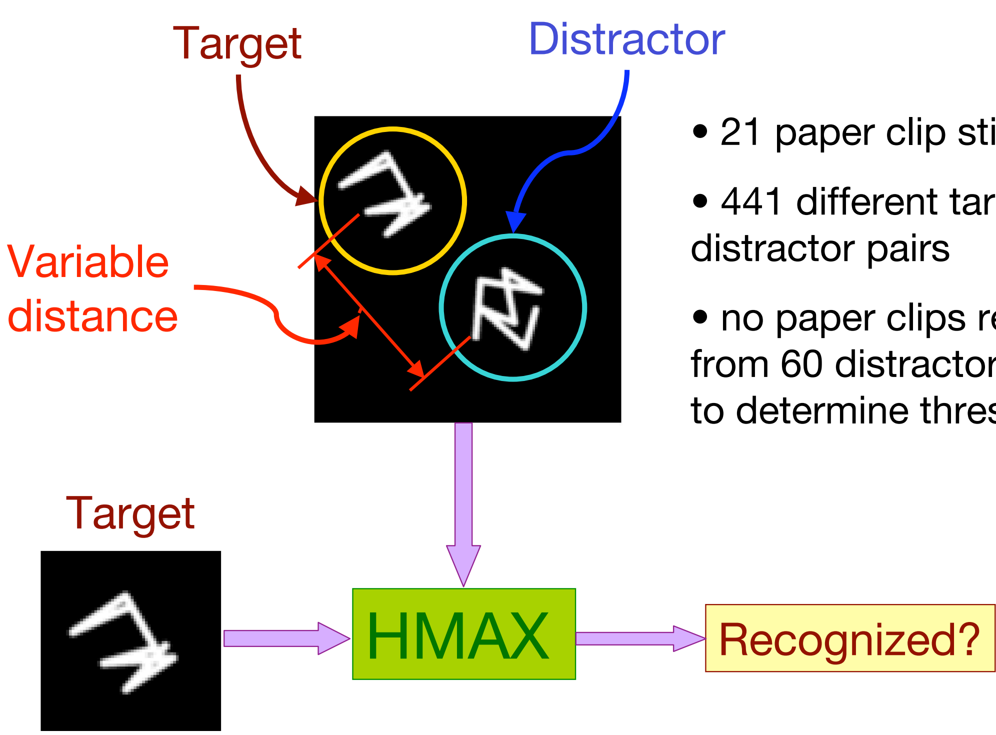


When is a stimulus recognized ?



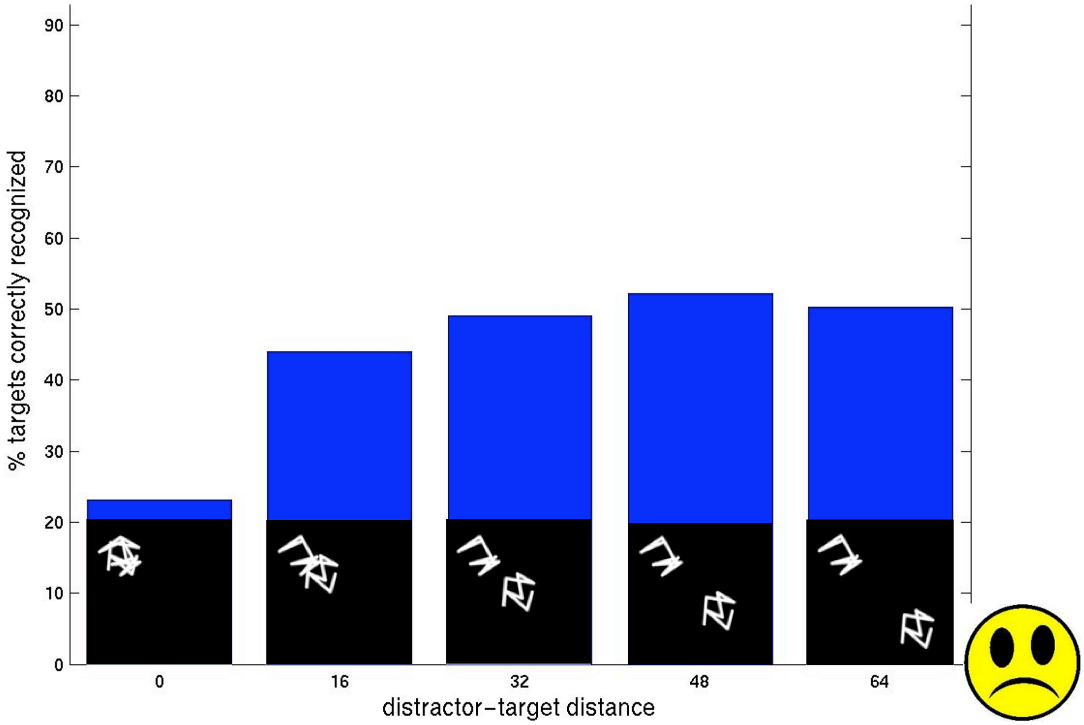
60 randomly chosen distracter paper clips

Two paper clips



- 21 paper clip stimuli
- 441 different target/ distractor pairs
- no paper clips re-used from 60 distractors used to determine threshold

Two paper clips - Results

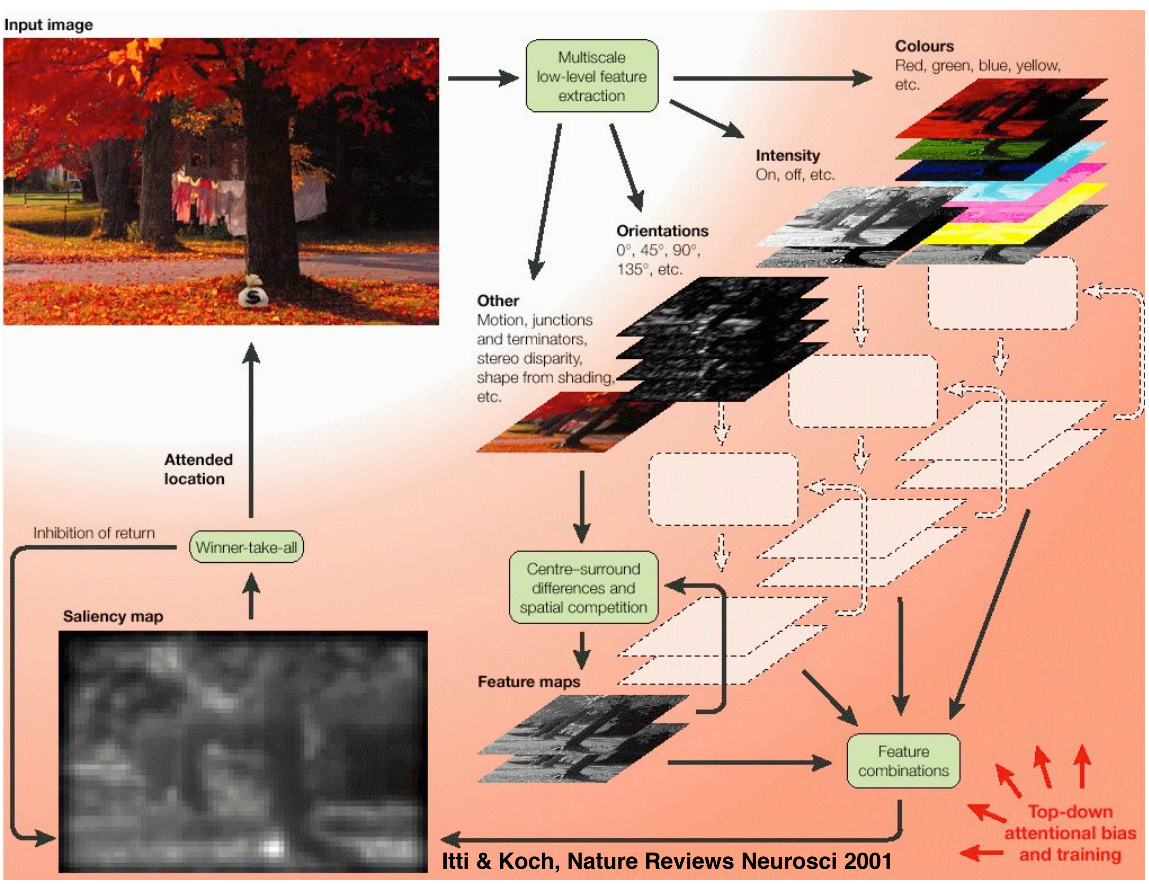


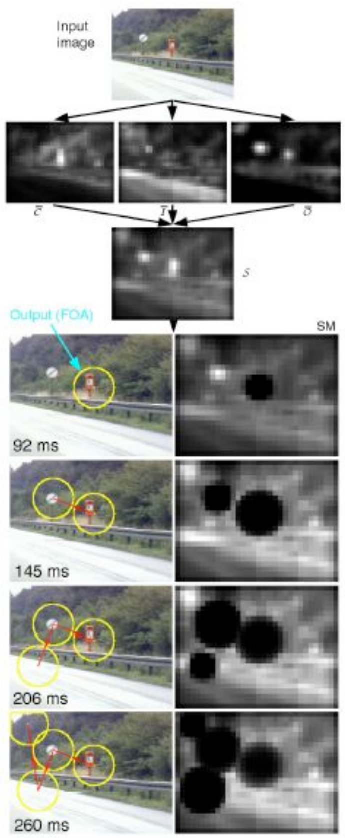
Hierarchical Template Matching for Object Recognition

- Image passed through layers of units with progressively more complex features at progressively less specific locations.
- Hierarchical in that features at one stage are built from features at earlier stages.
- Processing hierarchy yields activation of view-tuned units.
- A collection of view-tuned units is associated with one object.
- Object recognition is severely impaired in the presence of clutter
- At present, no learning algorithm for tuning the weights has been developed (but see Wersing and Körner 2003 and LeChun, 1998).

The saliency map model of attention

- Itti, L., Koch, C., Niebur, E. (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20:1254-1259.
- Itti, L., Koch, C. (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res., 40:1489-1506.





Example

Discussion

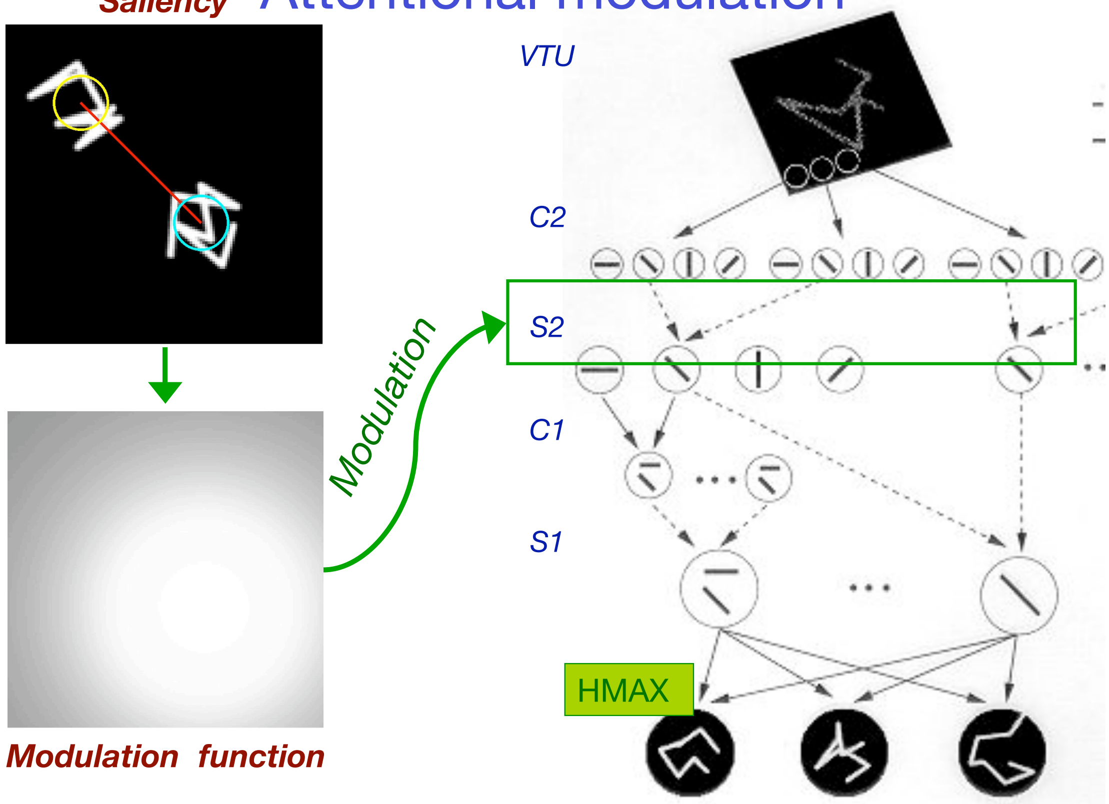
- The saliency model offers a fast algorithm for guiding vision to potentially meaningful parts of a scene.

- It selects only a point in space, as compared to an object or region. Region selection has to be added by a separate mechanism.
- Saliency is restricted to simple features.
- Attention is defined solely as the selection in space (no, or only indirect feature-based selection).
- The advantage of this mechanism for object recognition is limited, since a selection in space does not necessarily promote object-recognition.

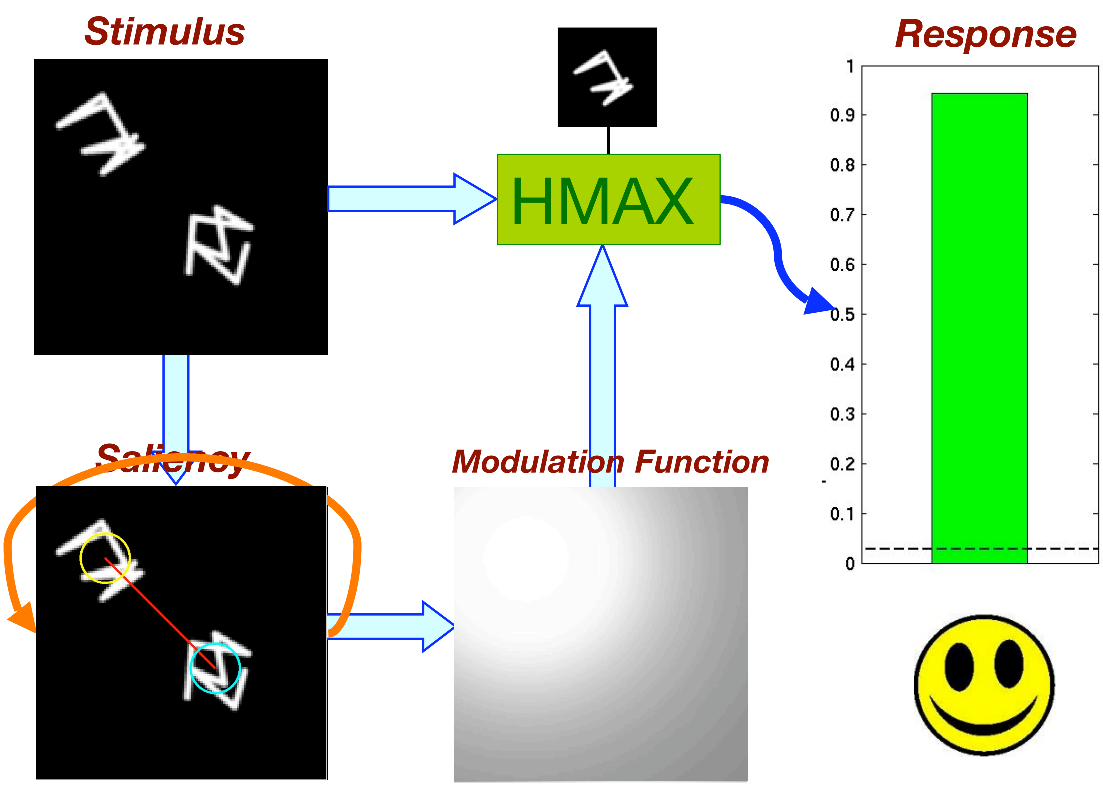
Combining the saliency map with hierarchical models of object recognition

- Walther, Itti, Riesenhuber, Poggio, Koch (2002) Attentional Selection for Object Recognition - a Gentle Way. In: Biologically Motivated Computer Vision. Lecture Notes in Computer Science. Berlin, Heidelberg, New York: Springer Verlag, 472-479.

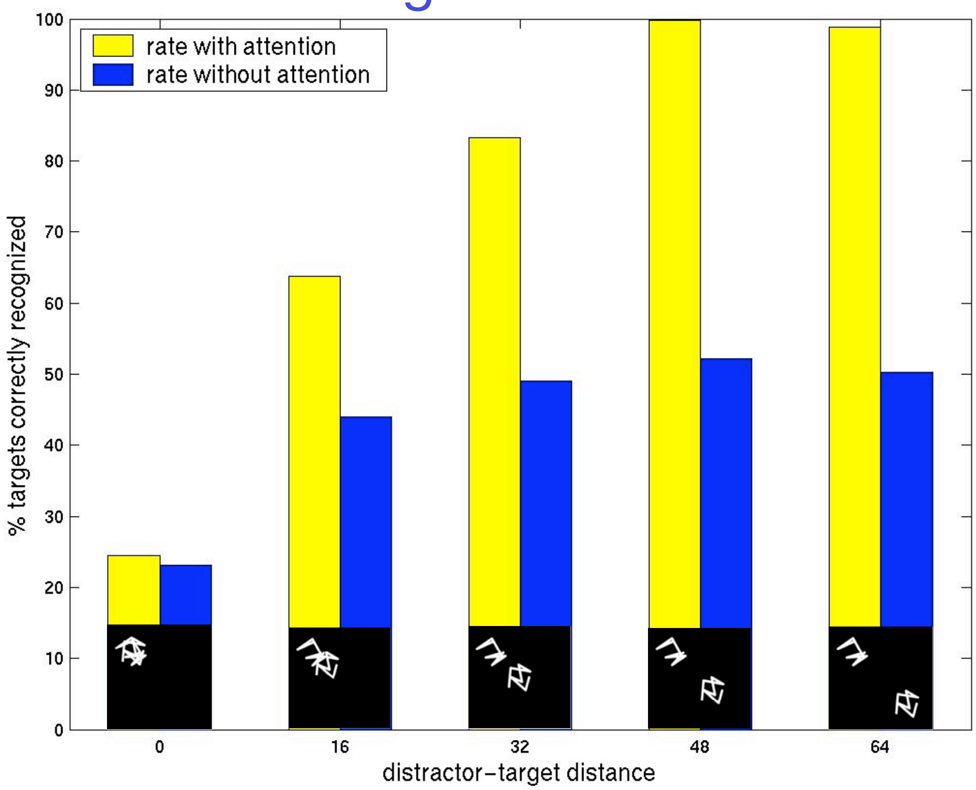
Attentional modulation



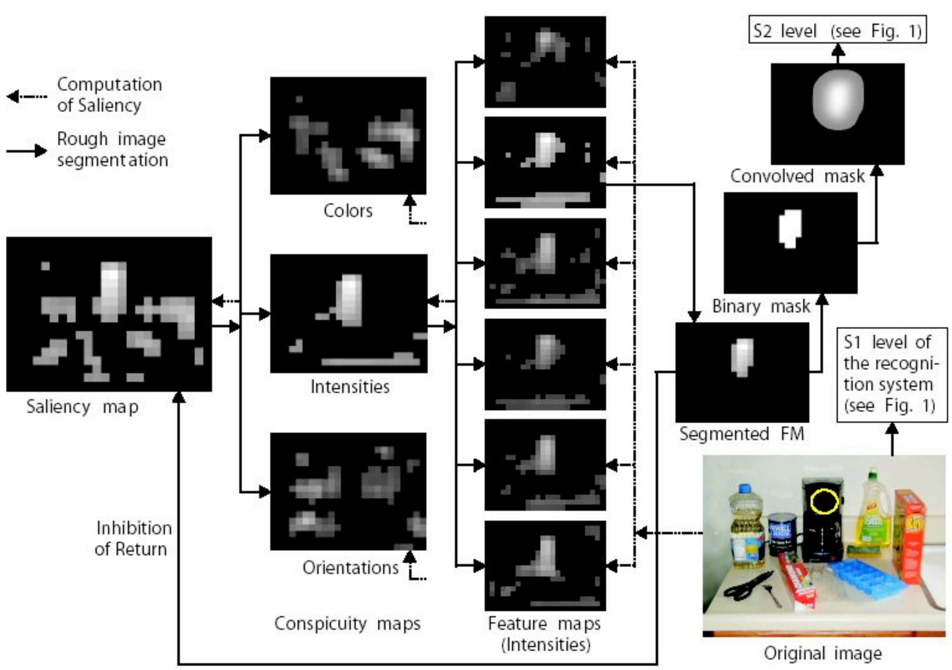
Concept of attention



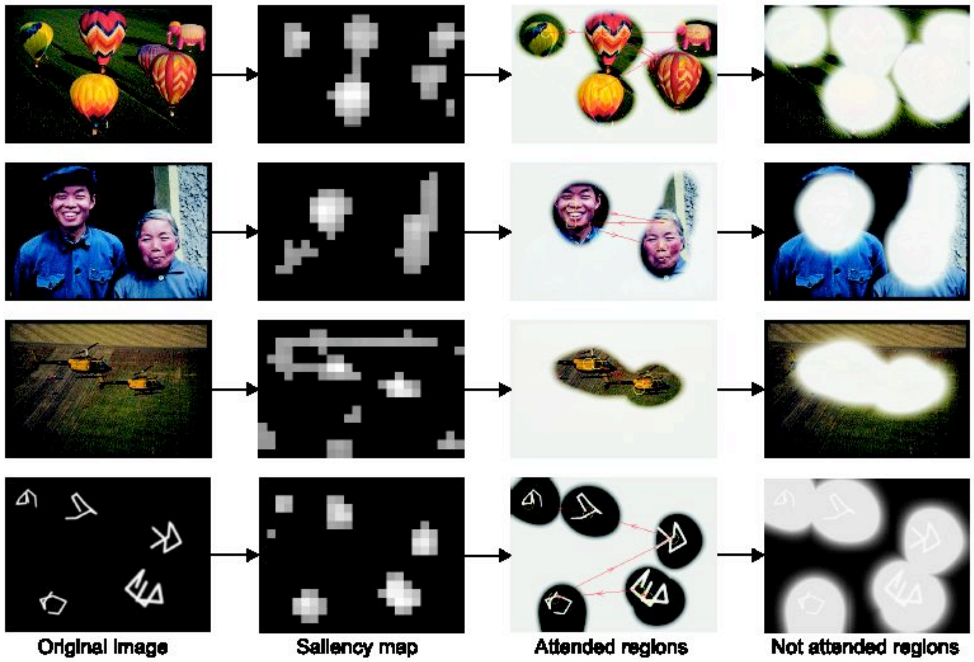
Recognition Rate



Towards attention and object recognition in natural scenes



Towards attention and object recognition in natural scenes

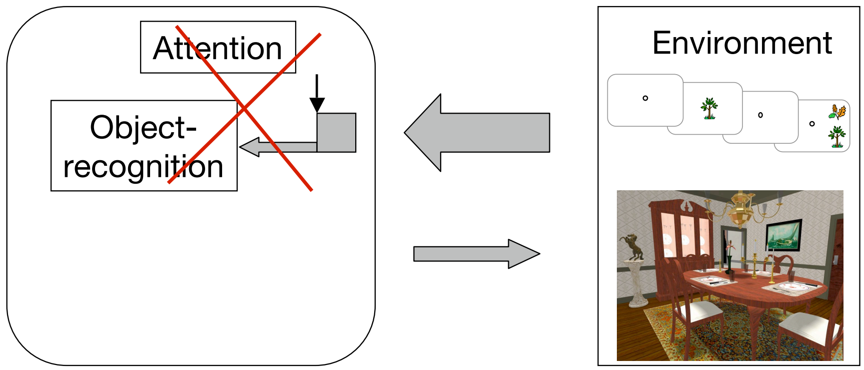


Discussion

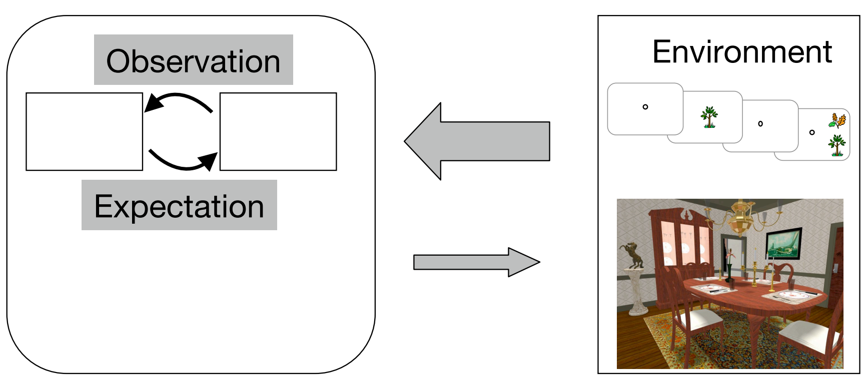
- The combination of the Saliency-Map model with a spatially selective focus and a hierarchical model for object recognition appears to be a straightforward way to go.

- Recognition depends on the quality of the focus.
- The focus is not determined by the recognition task.
- The model predicts that prior selection is necessary for object recognition, which appears to be a contradiction to the ability of category detection in dual-task situations.

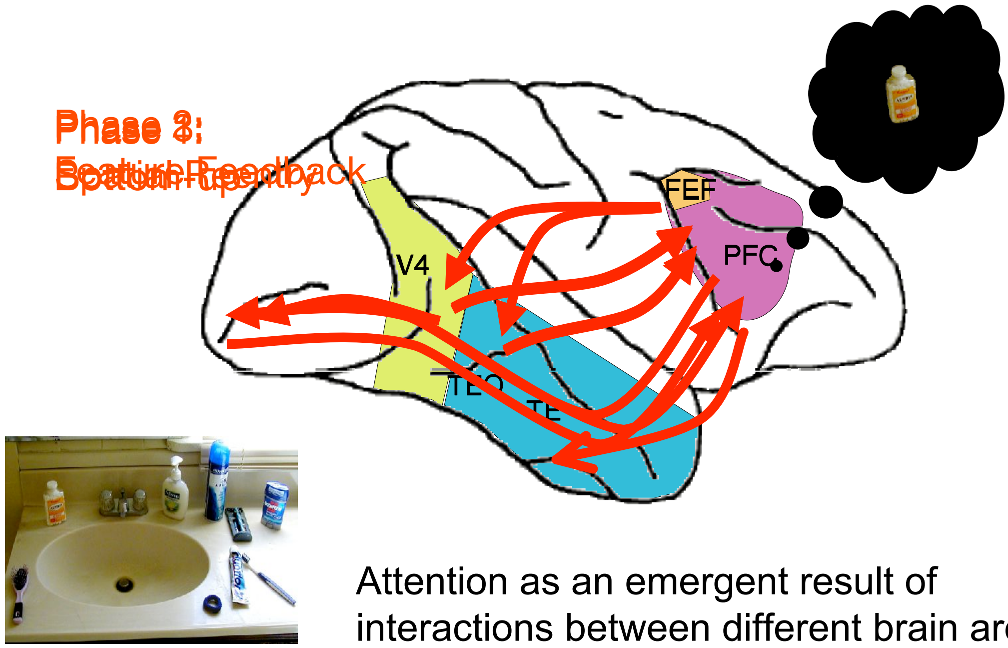
Classical approach of visual attention and object recognition



Alternative approach of visual attention and object recognition

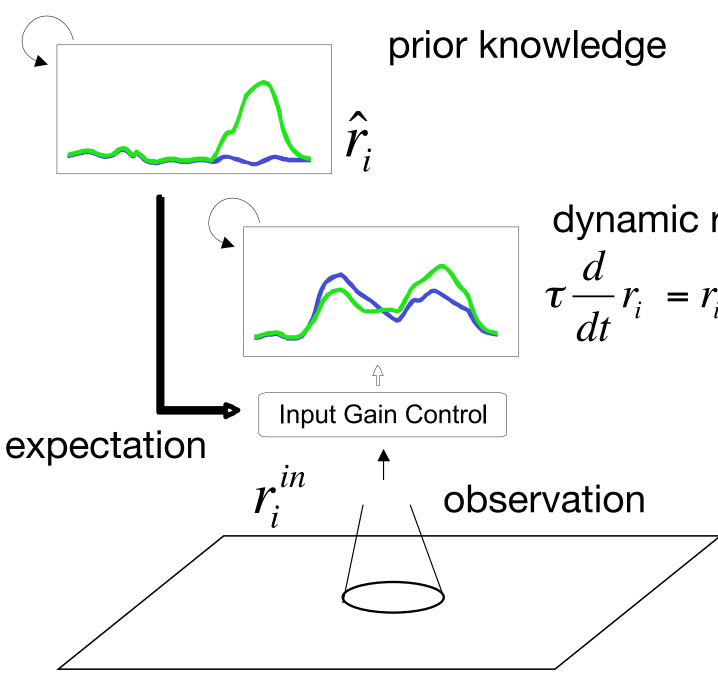


The three phases of visual perception



Attention as an emergent result of interactions between different brain areas!

The concept of population-based inference

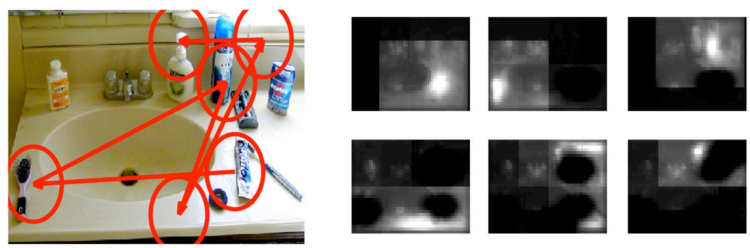


$$\tau \frac{d}{dt} r_i = r_i^{\text{in}} \left(1 + w \cdot \hat{r}_i\right) - r_i w_{inh} \sum_j r_j$$

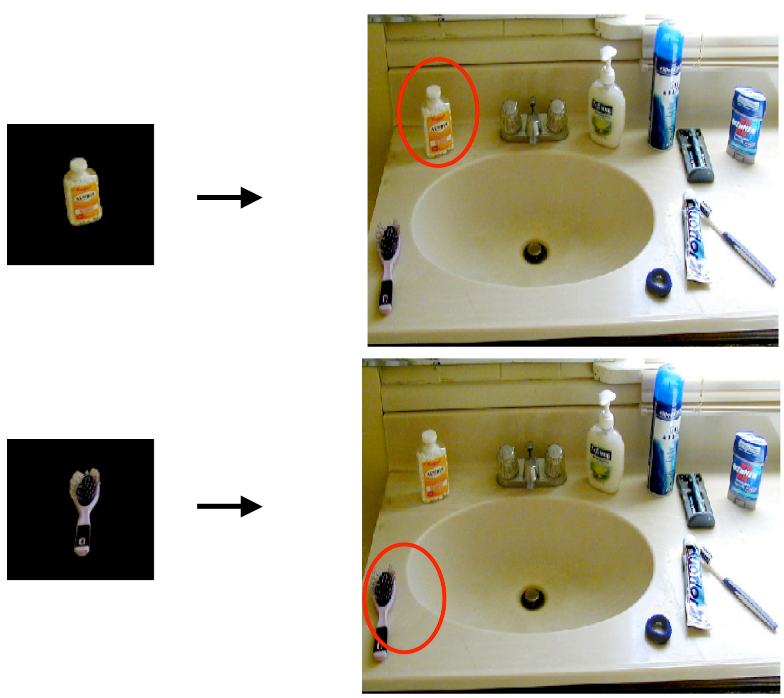
Search “without Target“

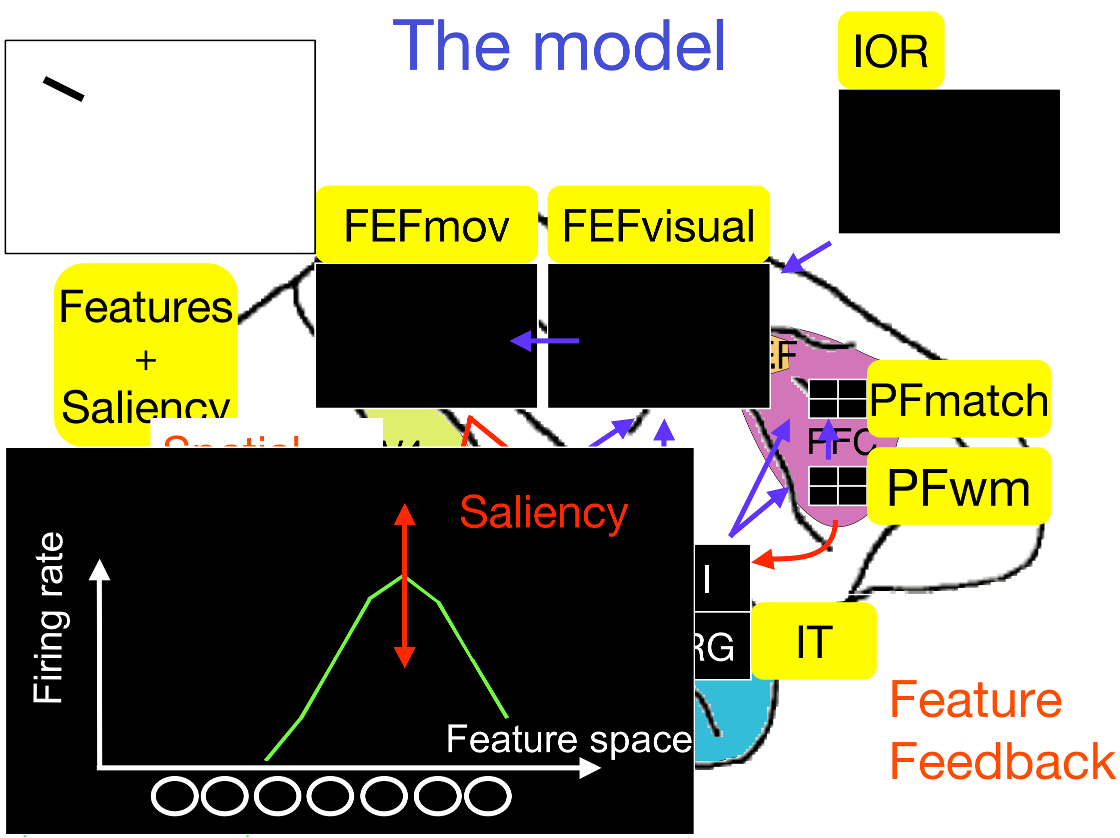
FEFv cells

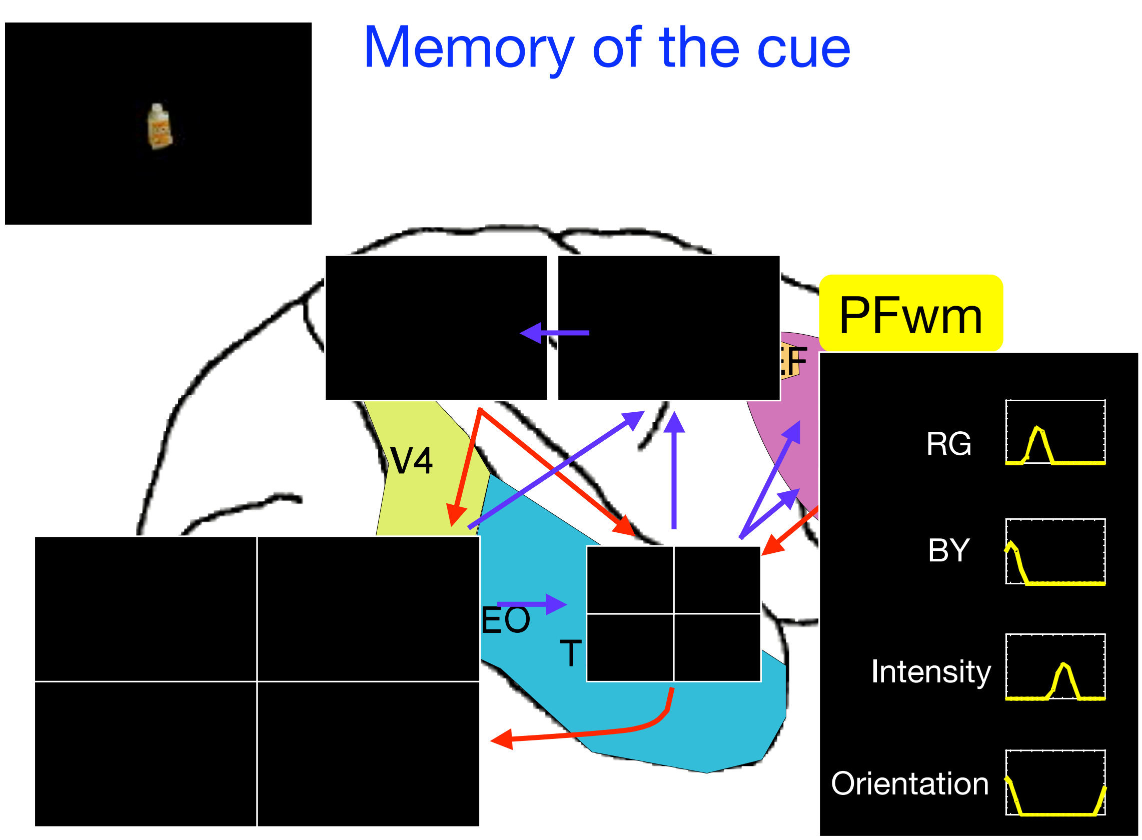
No Target



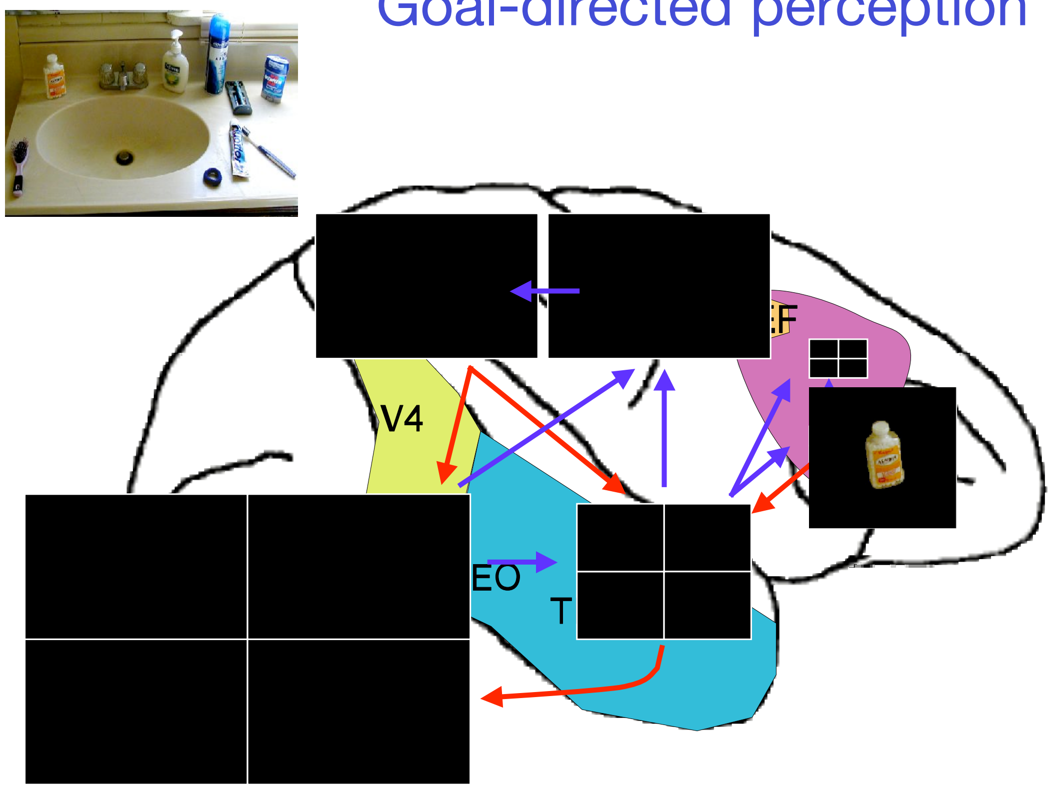
Search with Target

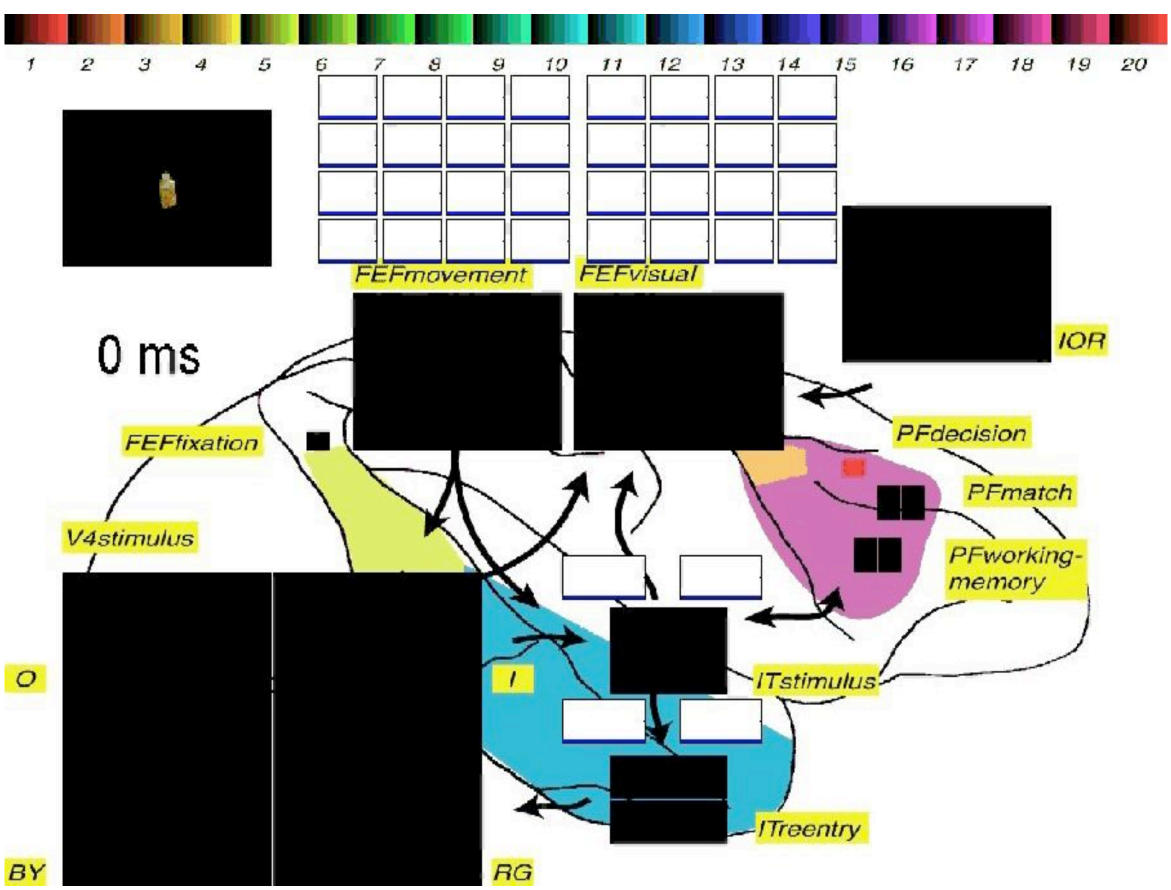



The model
IOR
FEFmov
FEFvisual
Features
+
Saliency
PFmatch
PFwm
Saliency
Firing rate
Feature space
IT
Feature
Feedback

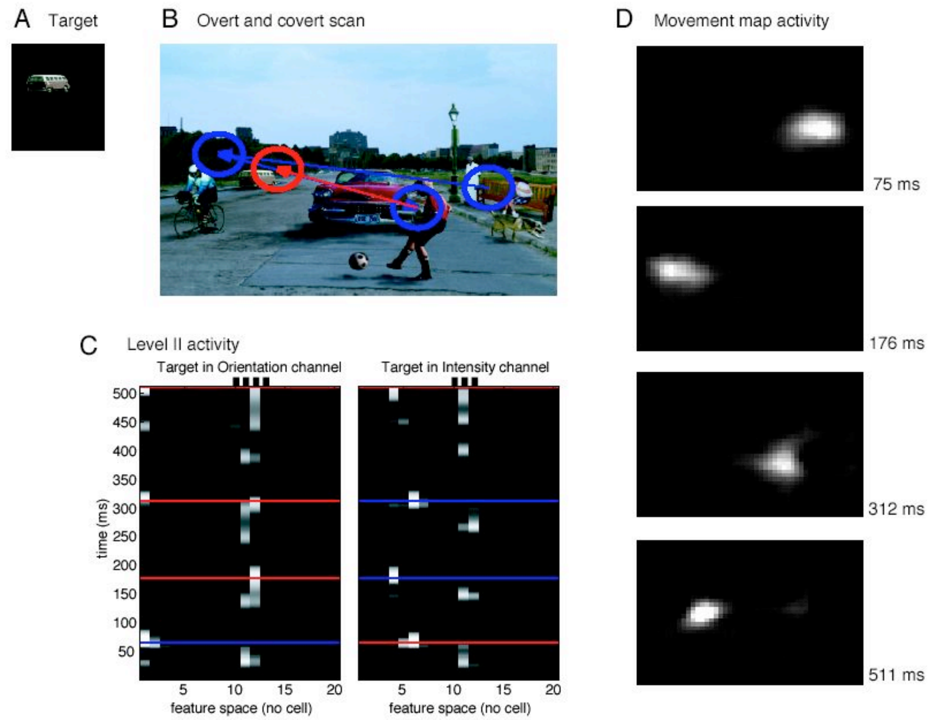

Memory of the cue
PFwm
V4
RG
BY
Intensity
Orientation

Goal-directed perception



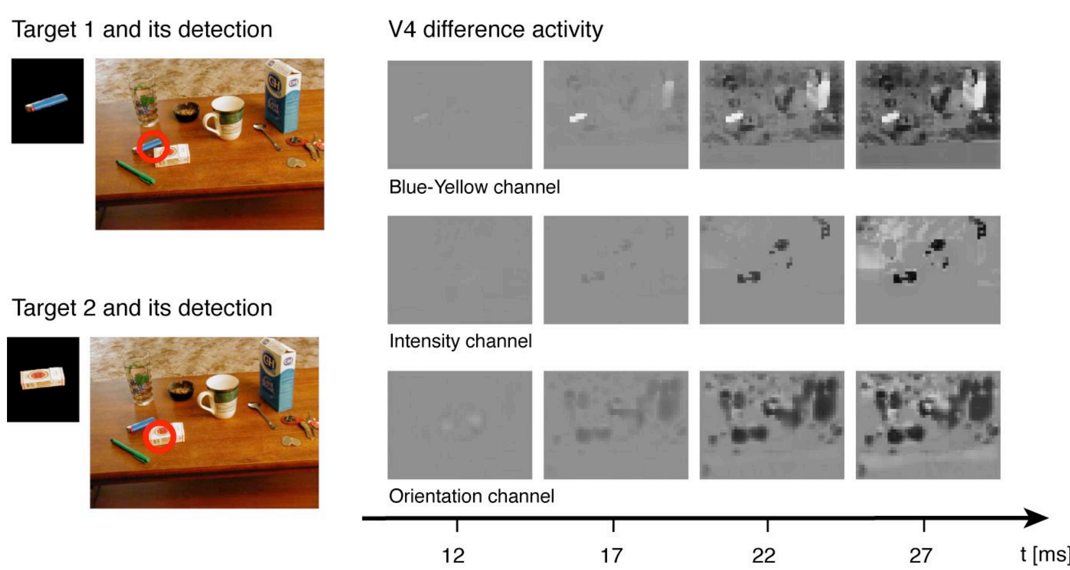


Overt and covert attention



Hamker, FH (2005) Journal for Computer Vision and Image Understanding.

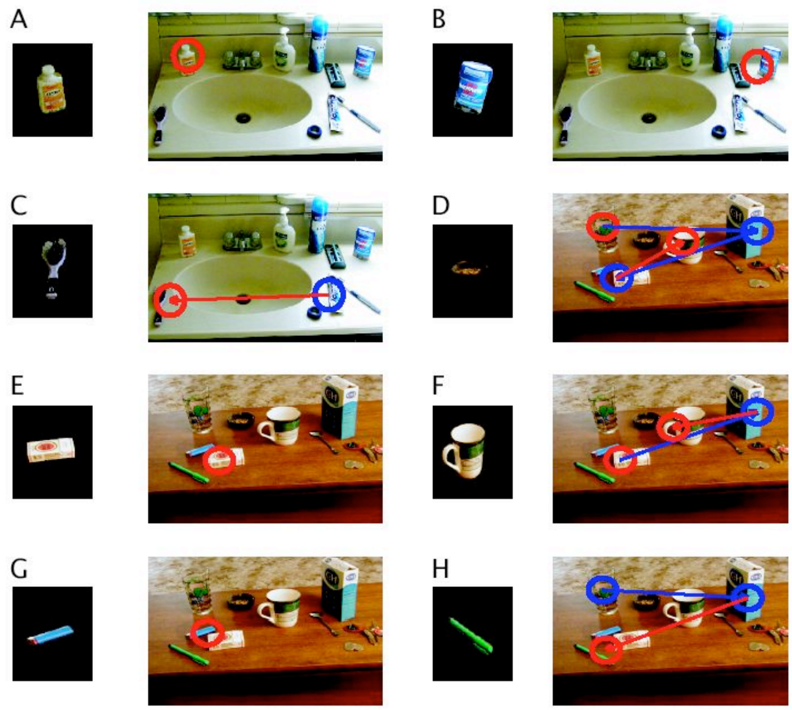
Feature-based attention in natural scenes



The model predicts that prior to any spatial selection, V4 contains information about potential target objects - feature-based attention.

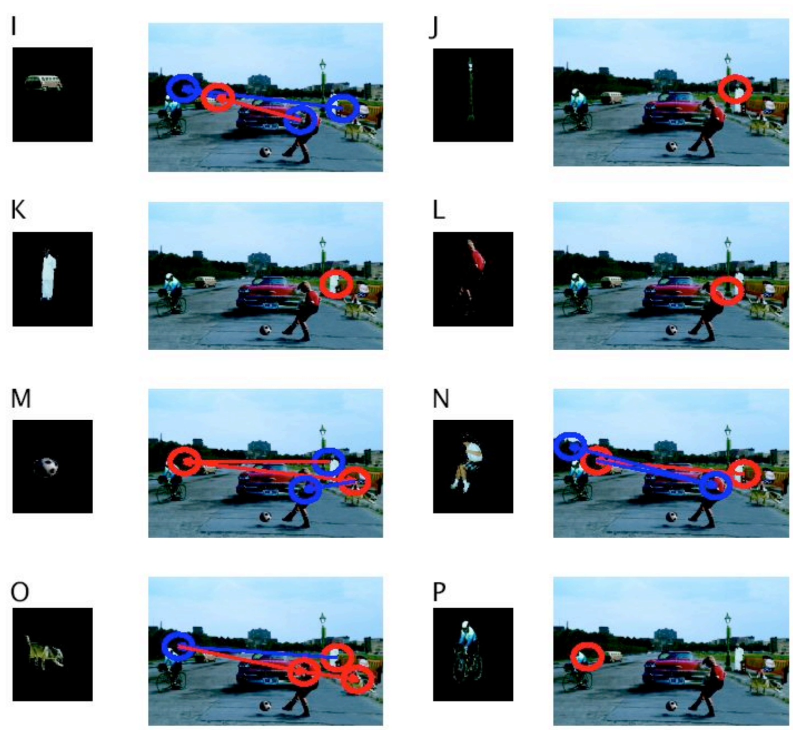
Hamker, F. H., BioSystems, 2006

Visual search examples



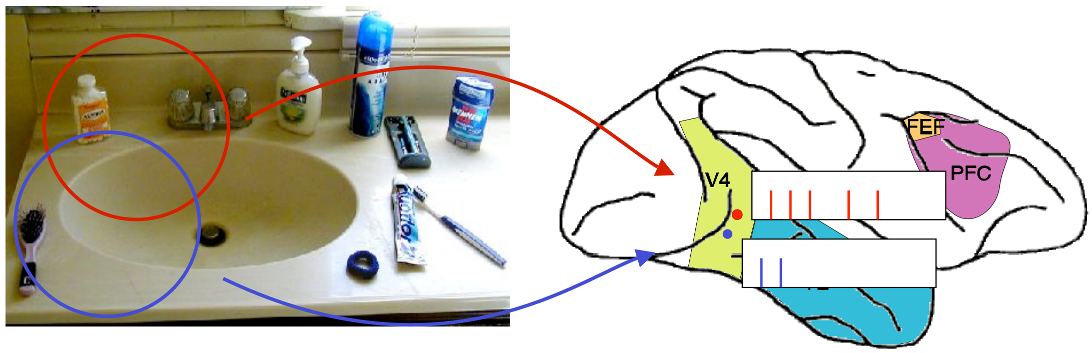
Hamker, FH (2005) Journal for Computer Vision and Image Understanding.

Visual search examples

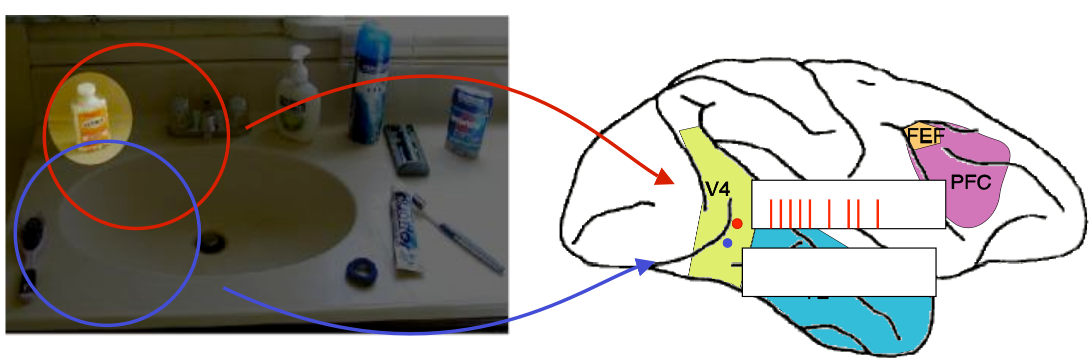


Hamker, FH (2005) Journal for Computer Vision and Image Understanding.

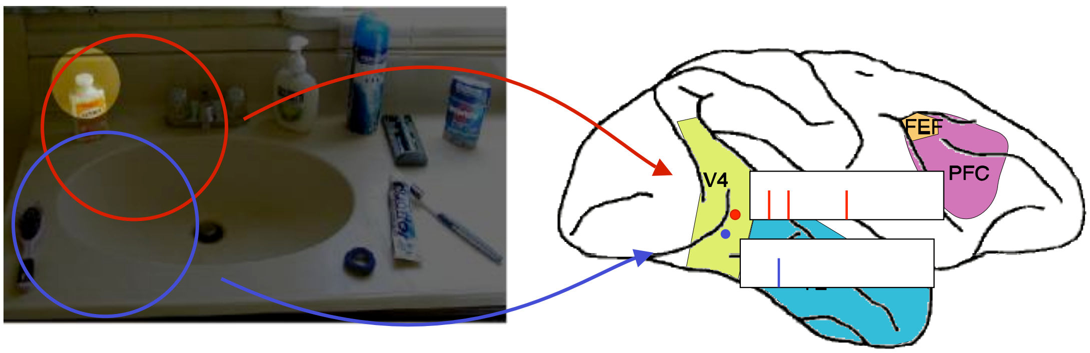
How does attention facilitate object recognition ?



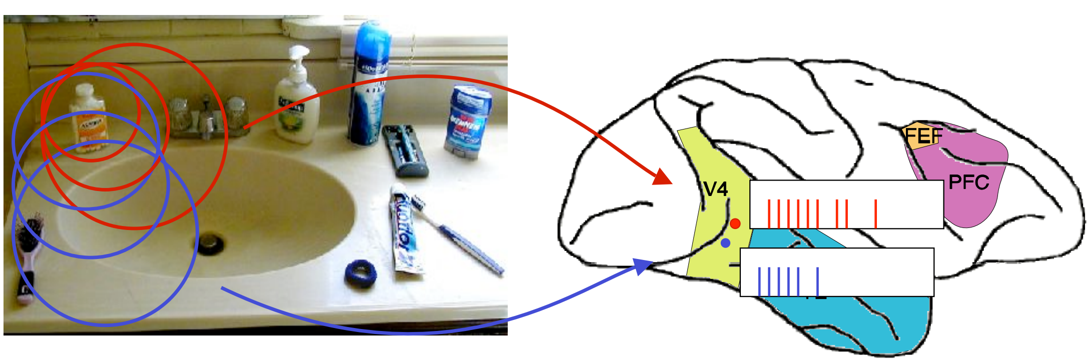
Spotlight Metaphor

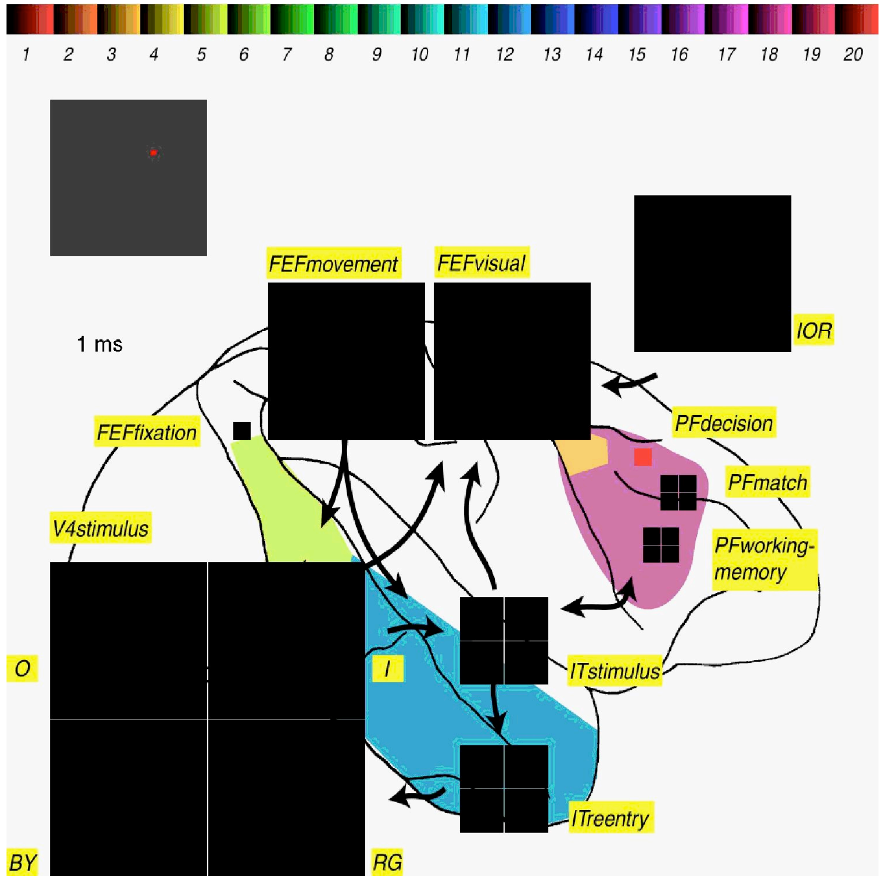


Problem of the Spotlight Metaphor

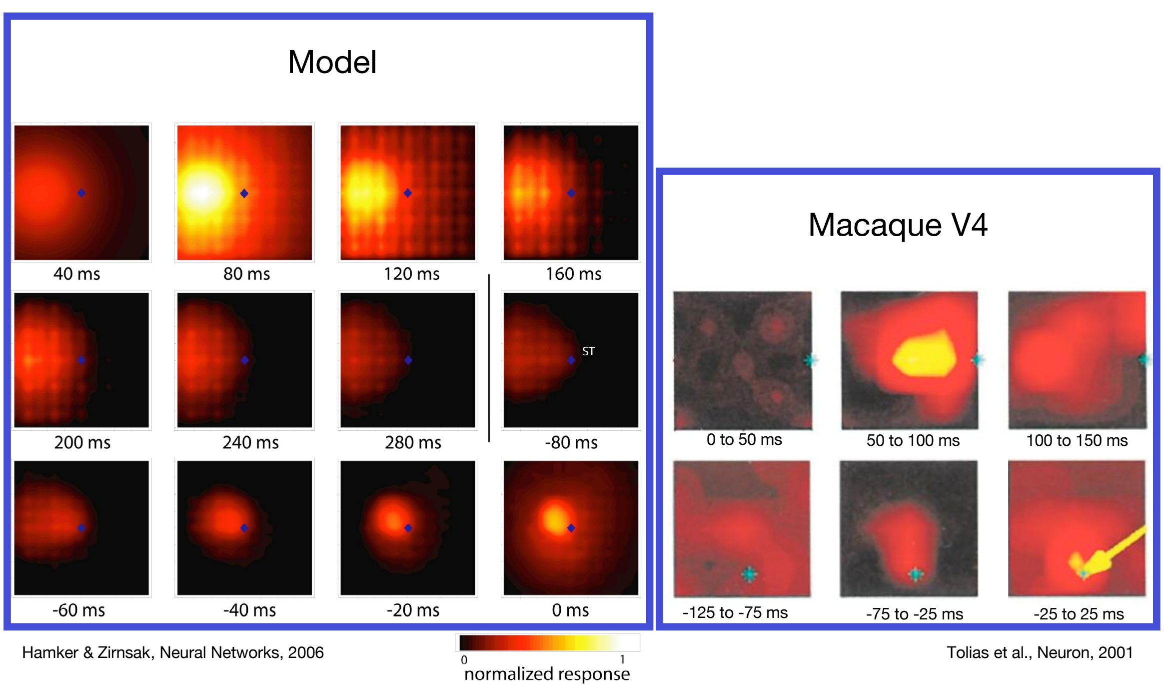


Attention tunes the RF properties





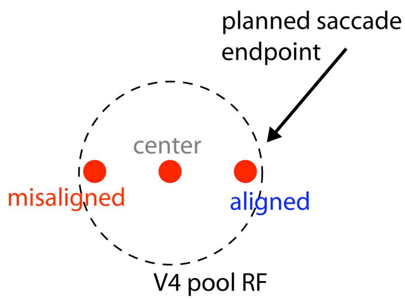
Model RF dynamics compared to V4

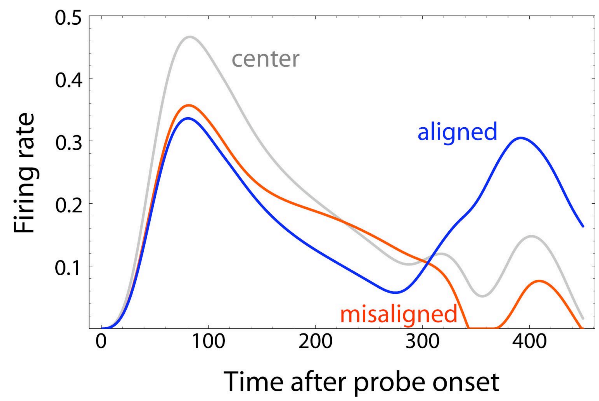


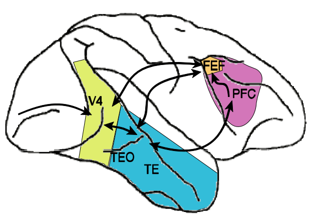
Hamker & Zirnsak, Neural Networks, 2006

Tolias et al., Neuron, 2001

Effects of a saccade on the neural firing rate of model V4 cells







Hypotheses

Attention is a network property

It emerges since high level task descriptions have to be connected to low level scene descriptions

The planning of an eye movement provides a reentry signal which influences perception

Feature-specific feedback within the object recognition pathway, gain control and competitive interactions directly enhance the features of interest and guide spatial attention to the object of interest.

I propose that the direction of attention and recognition must be an iterative process to be effective.

Limitations of the present approach

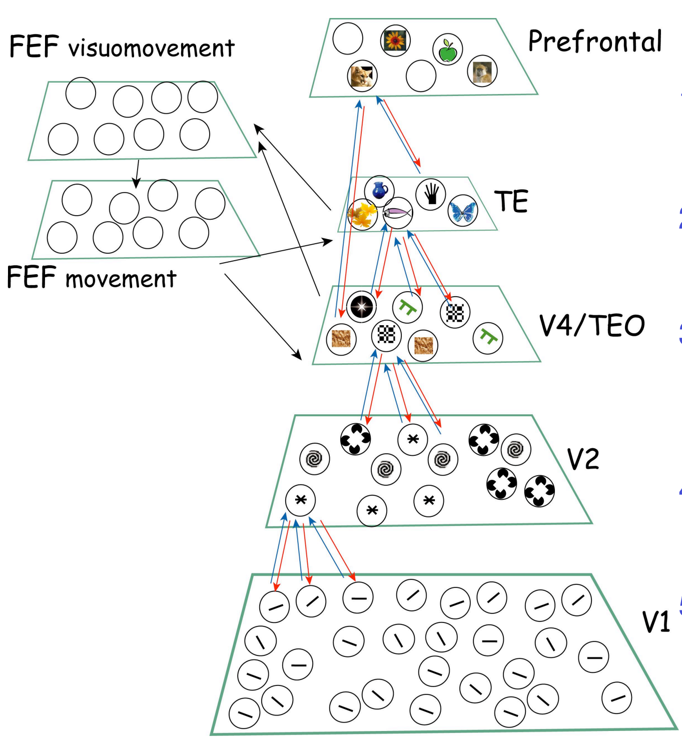
The representation on which object detection is made does hardly allow for real object recognition tasks and the guidance of vision can only be based on simple color and orientation cues.

Extend the present approach by learning feedback and feedforward transformations within the feature spaces of different complexity considering image statistics.

The model does not know what too look for.

Develop a computational approach to learn the recall of target features on the task at hand in a reward based scenario for guiding visual perception.

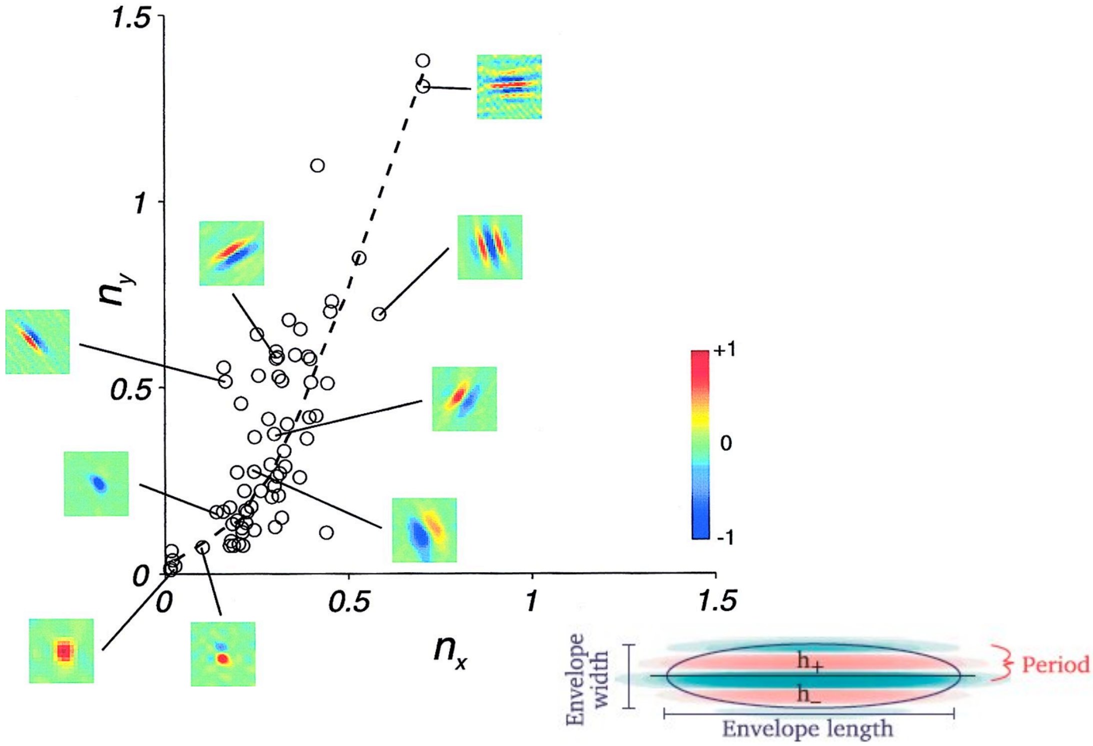
Vision as an active, constructive process



1. Top-down guidance – vision is guided by templates
2. Enhancement of features of interest – rather than only selecting just the location
3. Parallel pattern matching – switch into a serial search if parallel search does not discriminate the target
4. Fast bottom-up recognition – despite recurrent interactions in the system
5. Flexible pattern matching – matching process is flexible and indicates the similarity of target and object

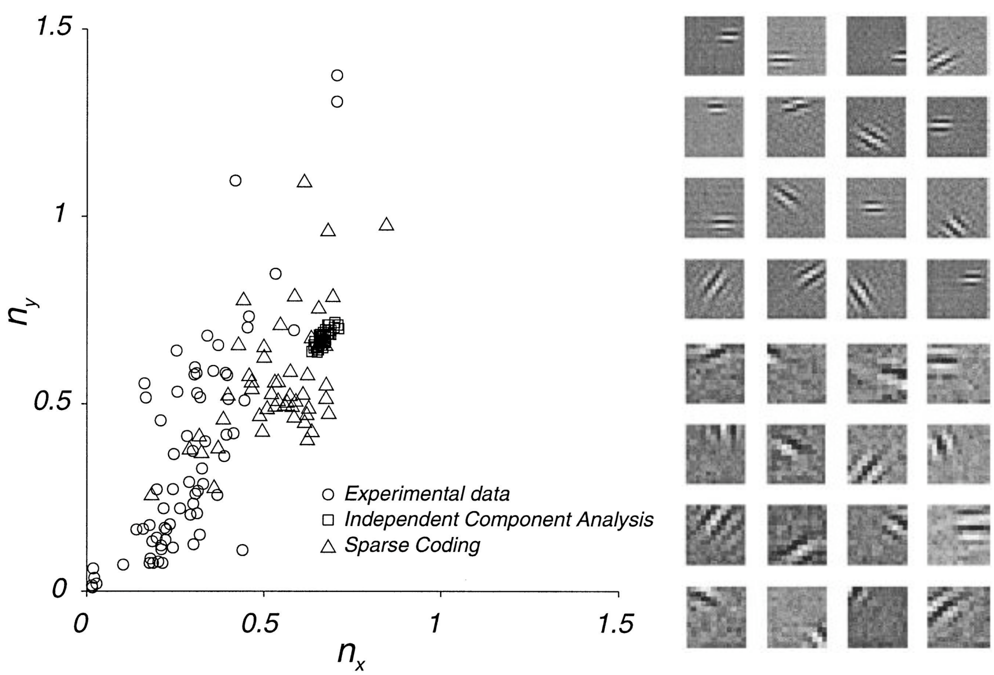
Data of V1 receptive fields

Ringach DL (2002) J Neurophysiol. 88:455-63.

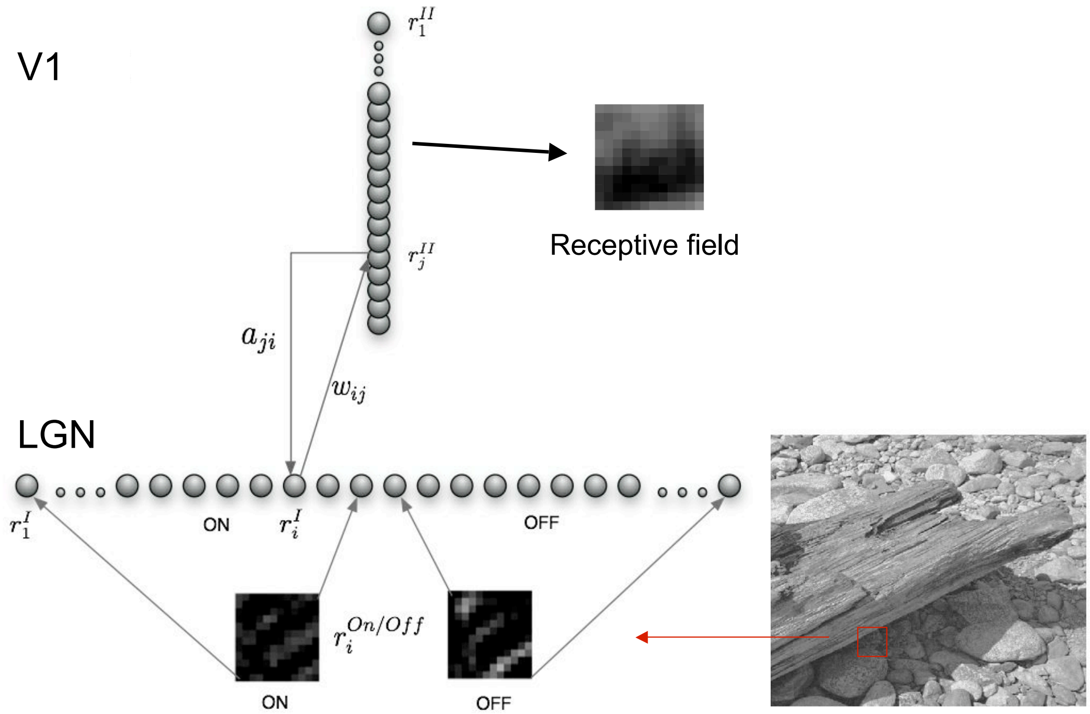


Models of V1 receptive fields

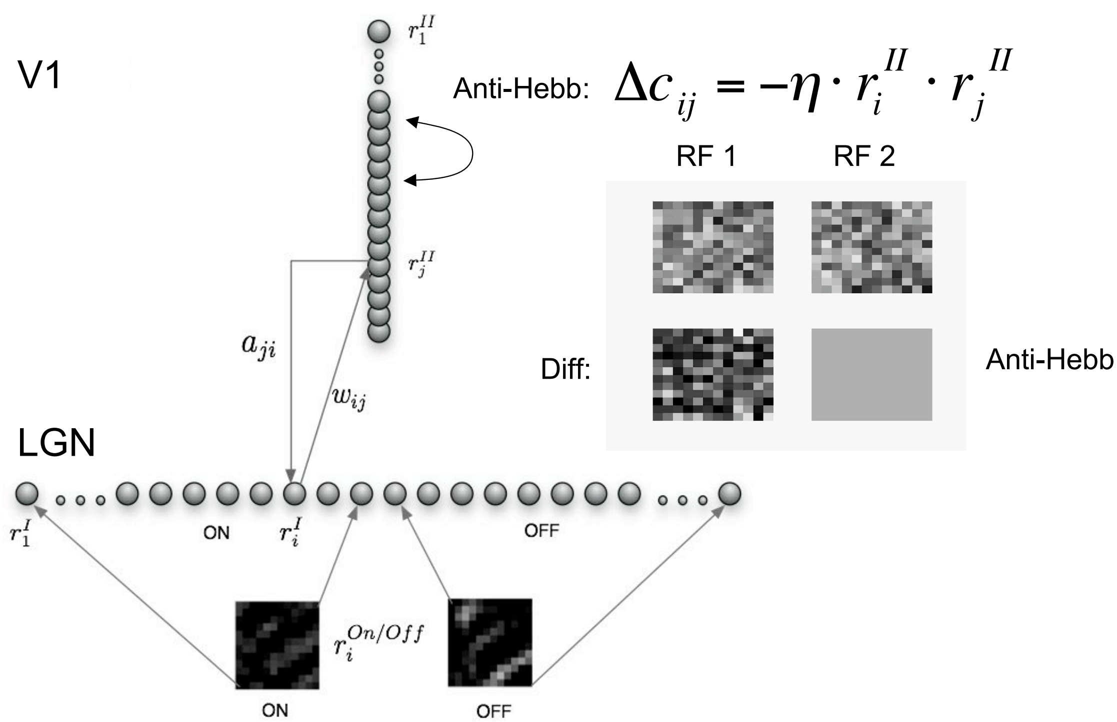
Ringach DL (2002) J Neurophysiol. 88:455-63.



Hebbian learning of RFs



Sparse coding by decorrelation



Results of learning in model V1

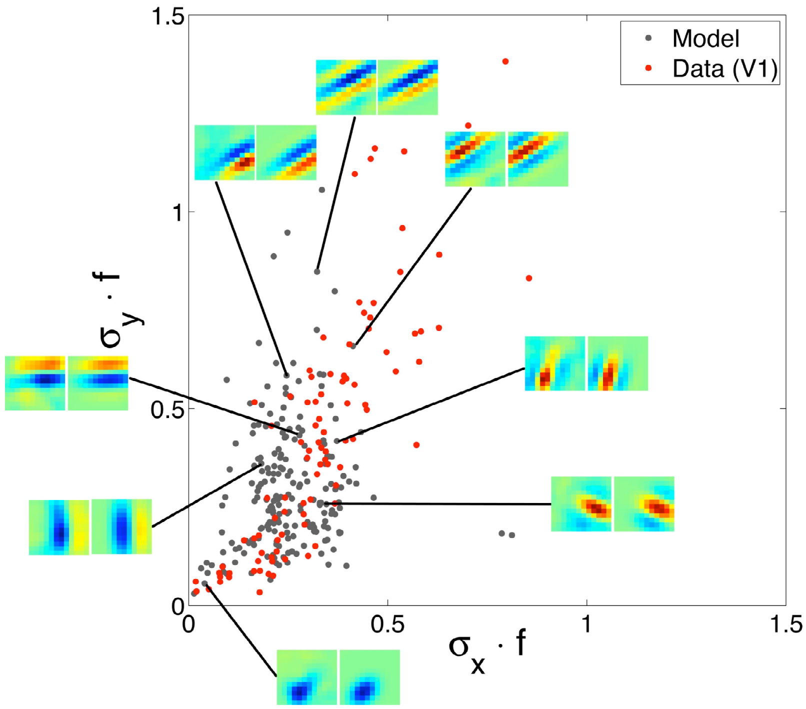
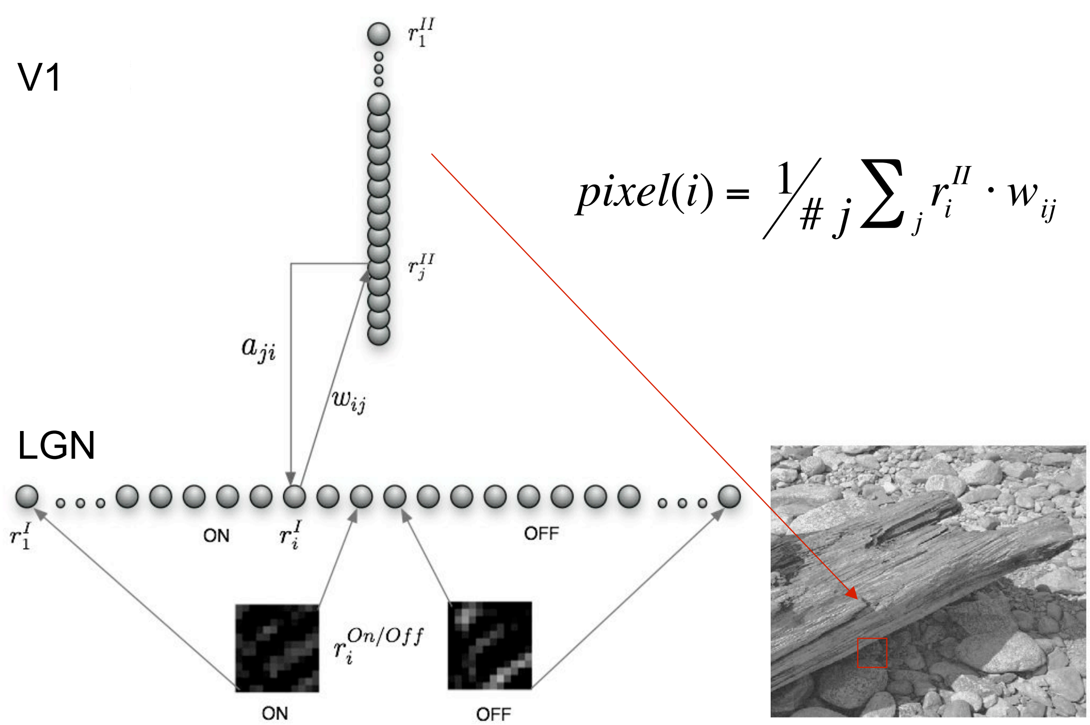
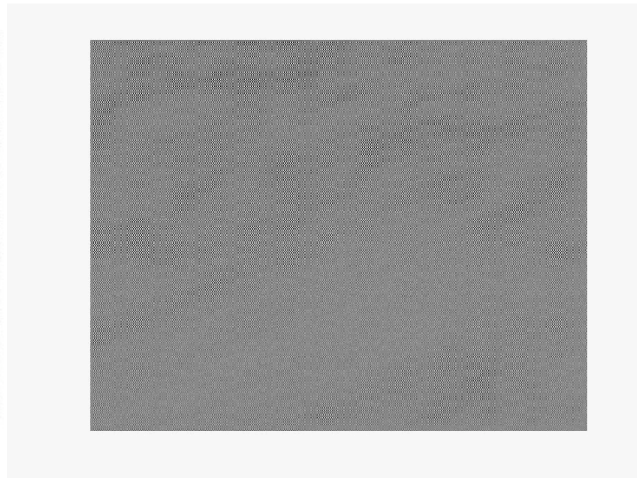


Image reconstruction



$$pixel(i) = \frac{1}{\# j} \sum_j r_i^{II} \cdot w_{ij}$$

Image reconstruction



Original | Reconstruction

Challenges and planned work in Eyeshots

- Learning of joint feature and disparity information in model V1 receptive fields, by training the model with stereo images (requires images sequences).
- Expand approach to the next higher level to learn more complex features (including disparity).
- Implement attentional dynamics within this network.
- Learn when to look for a particular object through reward.

Deliverables in Eyeshots

Deliverables

D3.1a: Demonstration of learning disparity-tuned feature selective cells. Software module. (Month 12).

D3.1b: Demonstration of object selective cells at intermediate complexity showing properties of disparity. Technical report. (Month 24)

D3.2: Object-based top-down selection using learned bi-directional connections between feature detectors to localize the object of interest in a cluttered 3D scene. Software module. (Month 36)

D3.3a: A model of working memory that allows to activate context information for the task at hand based on the association of previous events leading to reward. We will use the learned feature responses (WP2) on real world scenes if the feature-detectors are available, otherwise artificial representations will be used. Software module. (Month 24).

D3.3b: Final, fully tested version of the Working Memory Model. Technical report. (Month 36).