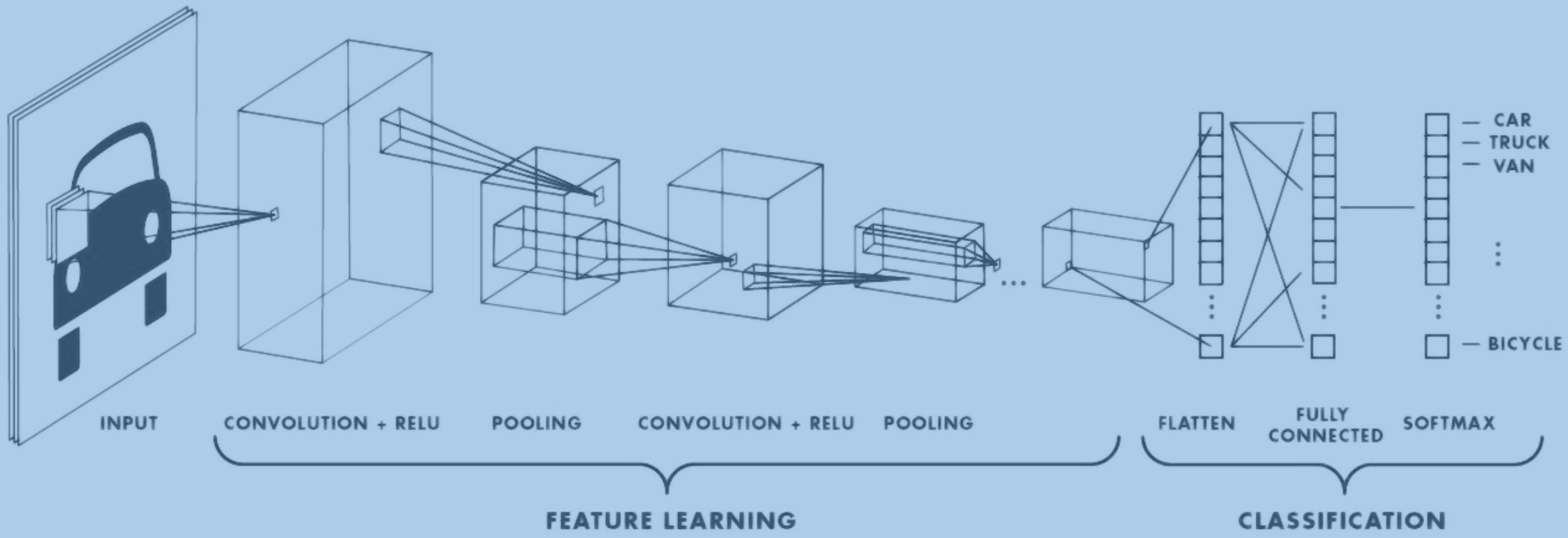
Bridging DFT and DNNs:

A neural dynamic process model of scene representation, guided visual search and scene grammar in natural scenes

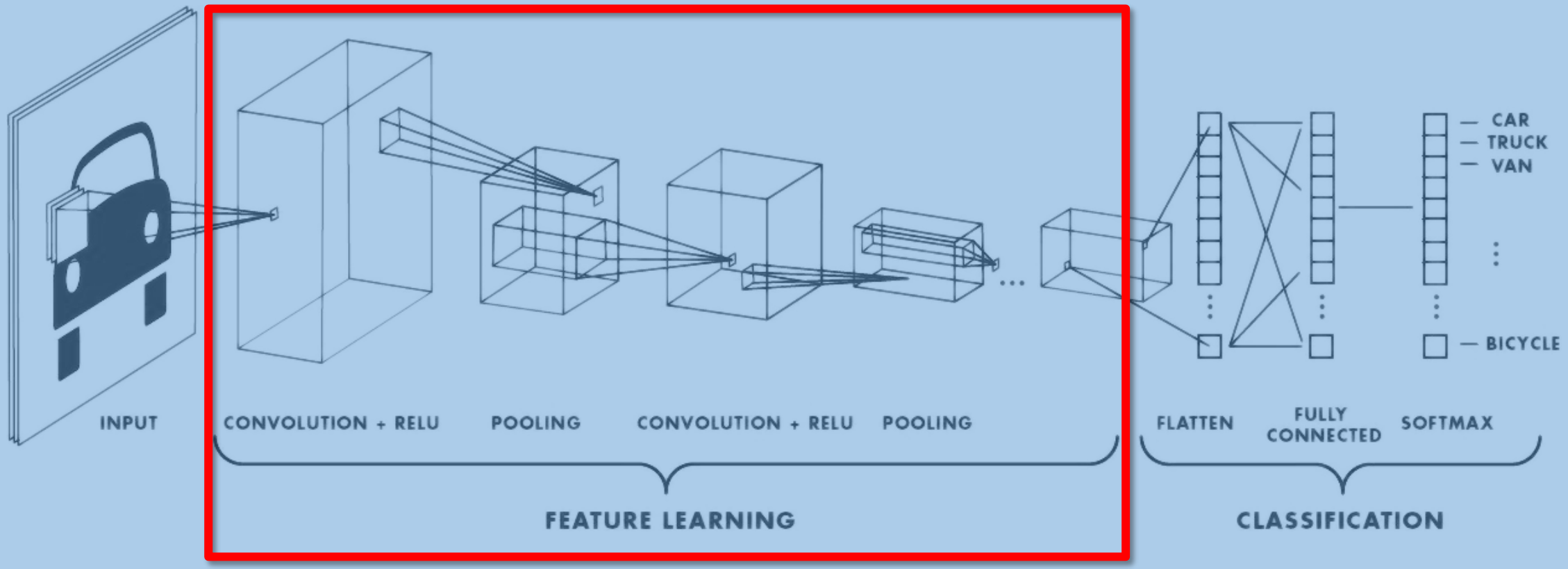
Raul Grieben

18.08.2022

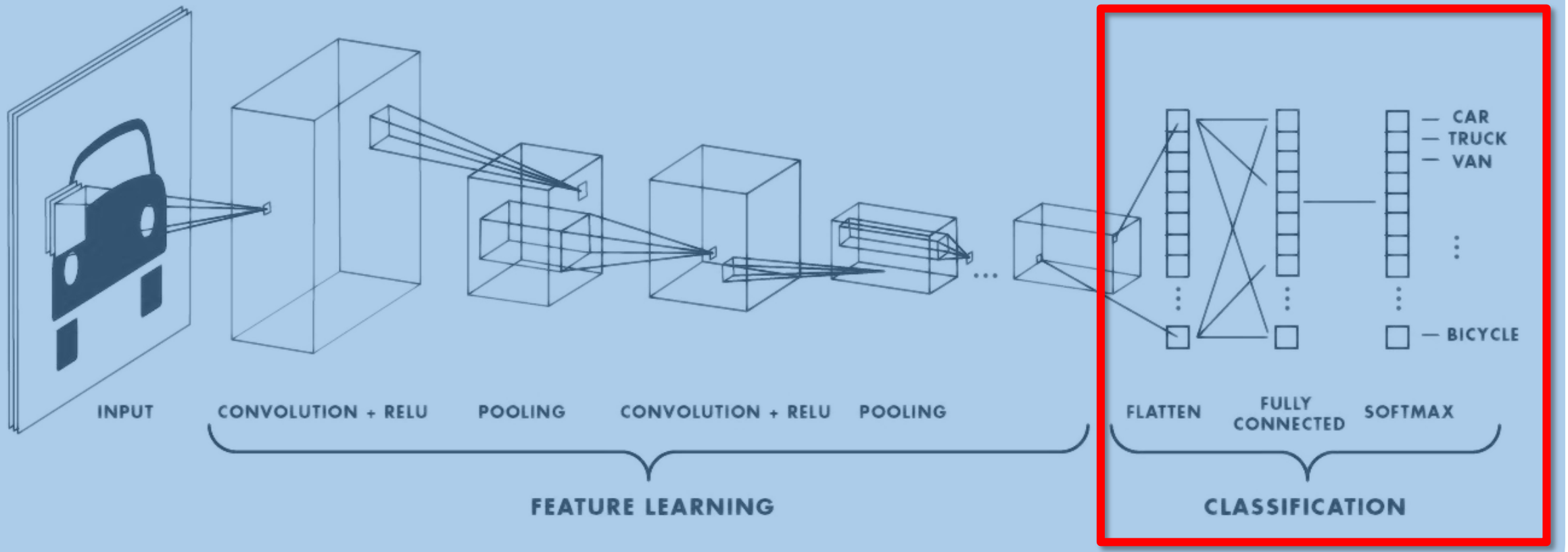
Deep CNN


CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION

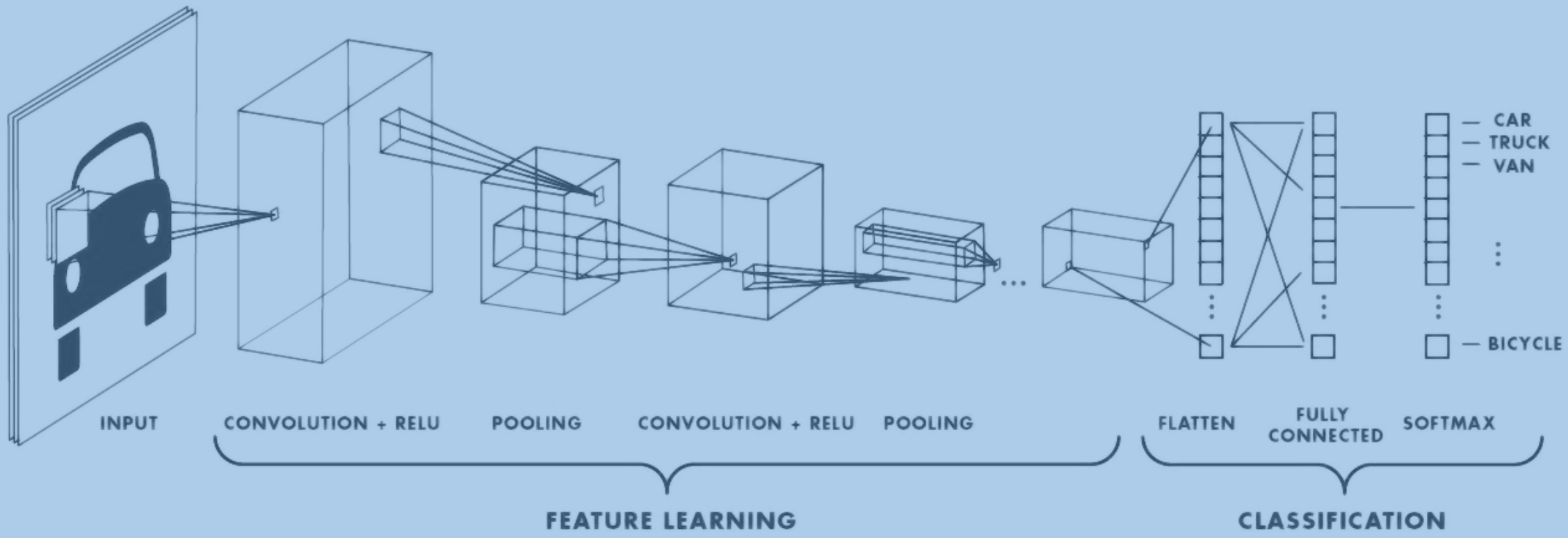
Deep CNN


INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FEATURE LEARNING
FLATTEN
FULLY CONNECTED
SOFTMAX
CLASSIFICATION
CAR
TRUCK
VAN
BICYCLE

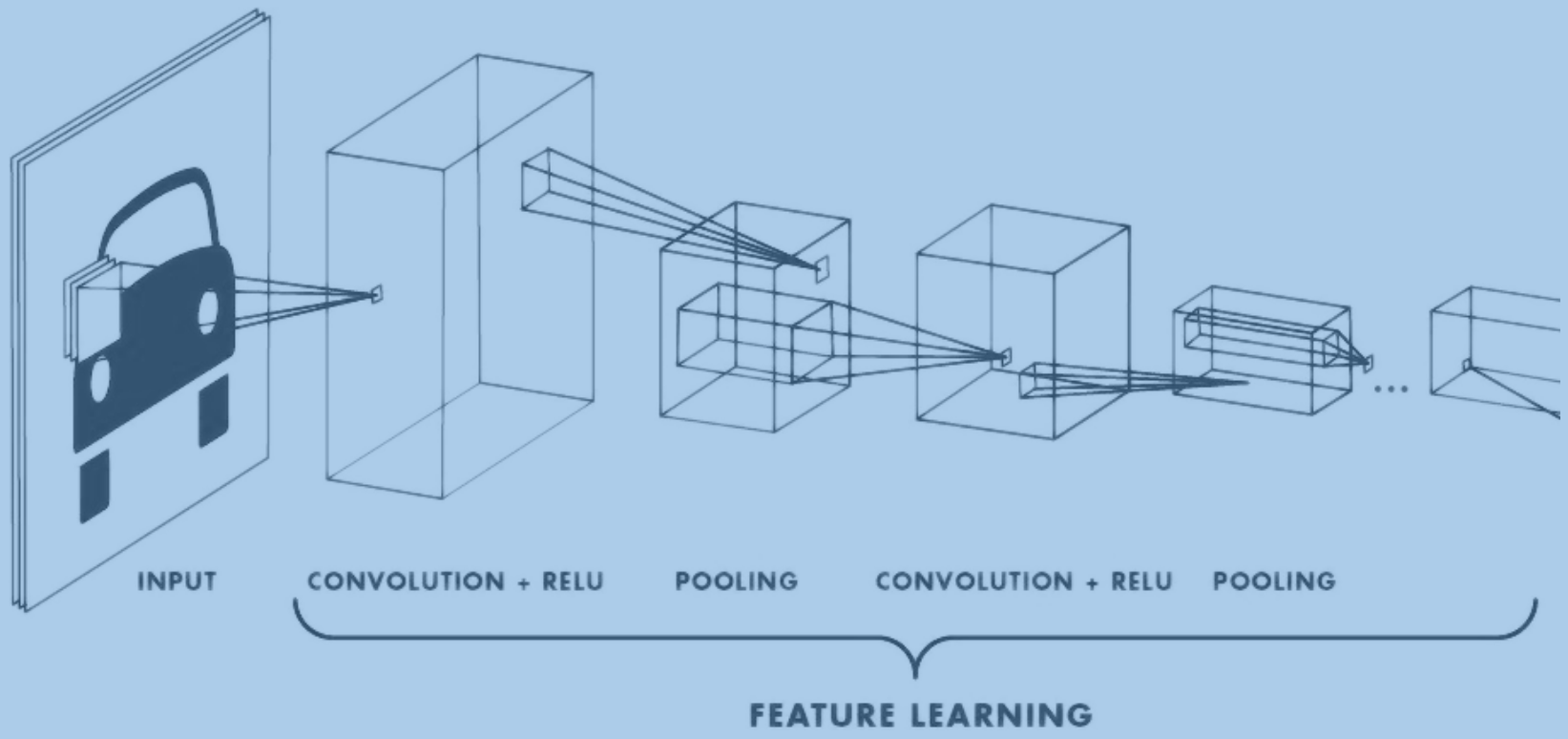
Deep CNN


CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION

Deep CNN

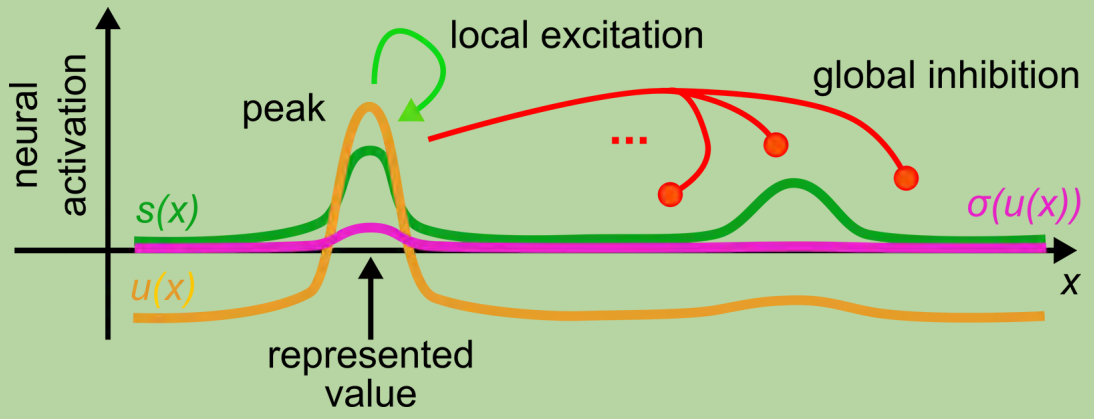

CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION

Deep CNN



Feature Maps

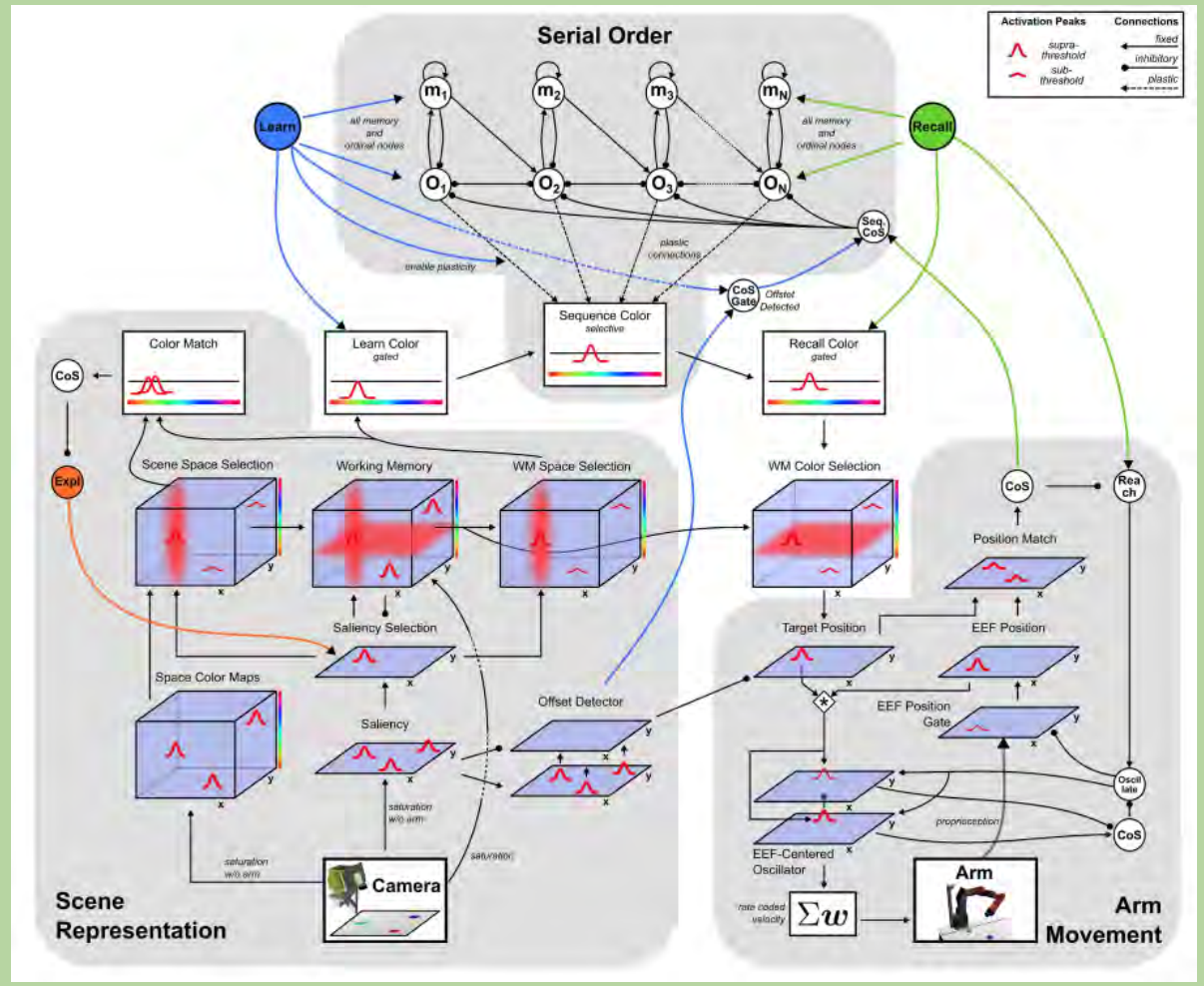
Dynamic Field Theory



$$\tau \dot{u}(x,t) = -u(x,t) + h + s(x,t) + \xi(x,t) + \int \omega(x-x')\,\sigma(u(x',t))\,dx'$$

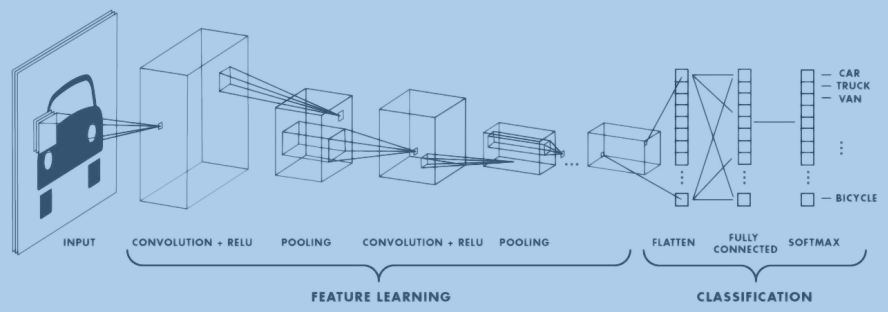
Schöner, G., Spencer, J. P., & The DFT Research Group (2016). *Dynamic thinking: A primer on dynamic field theory*. Oxford University Press

Dynamic Field Theory



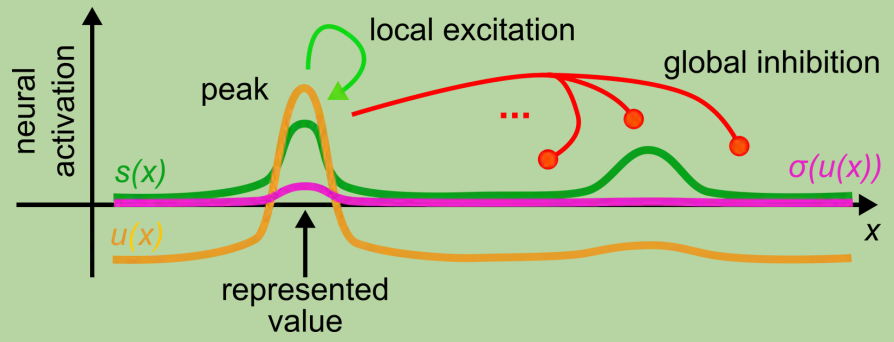
Tekülve, J., Fois, A., Sandamirskaya, Y., & Schöner, G.. (2019). Autonomous Sequence Generation for a Neural Dynamic Robot. Frontiers in Neurorobotics

Deep CNN



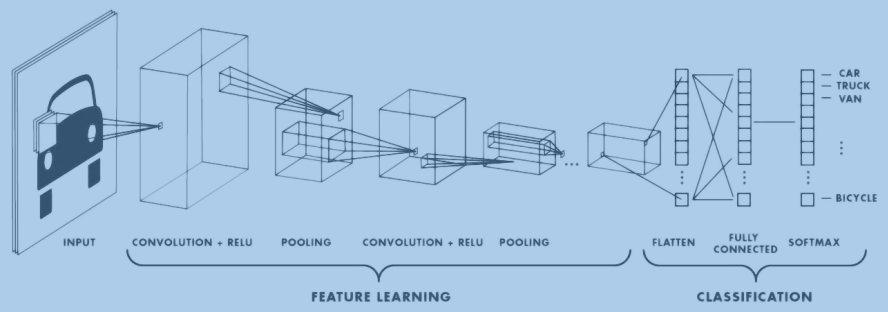
Feed-forward path

Dynamic Field Theory



Higher cognition

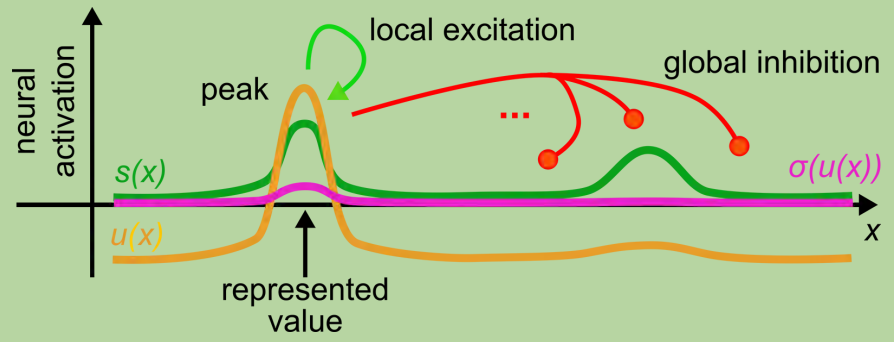
Deep CNN



Feed-forward path

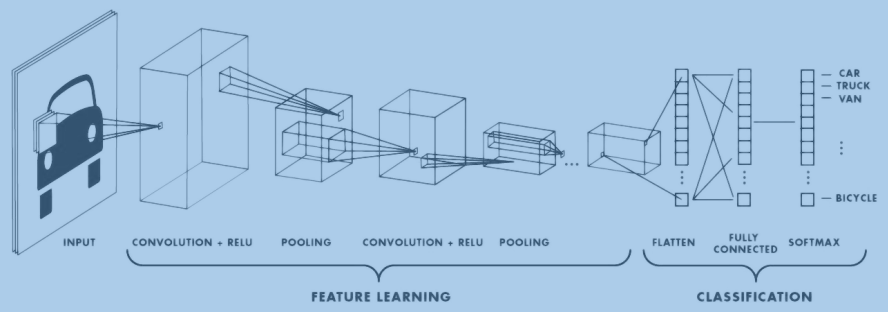
interface

Dynamic Field Theory



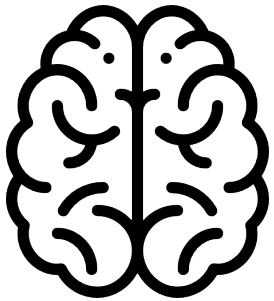
Higher cognition

Deep CNN

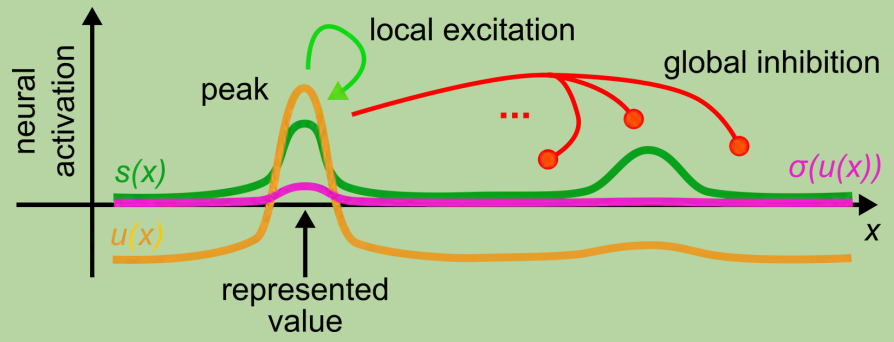


Feed-forward path

interface

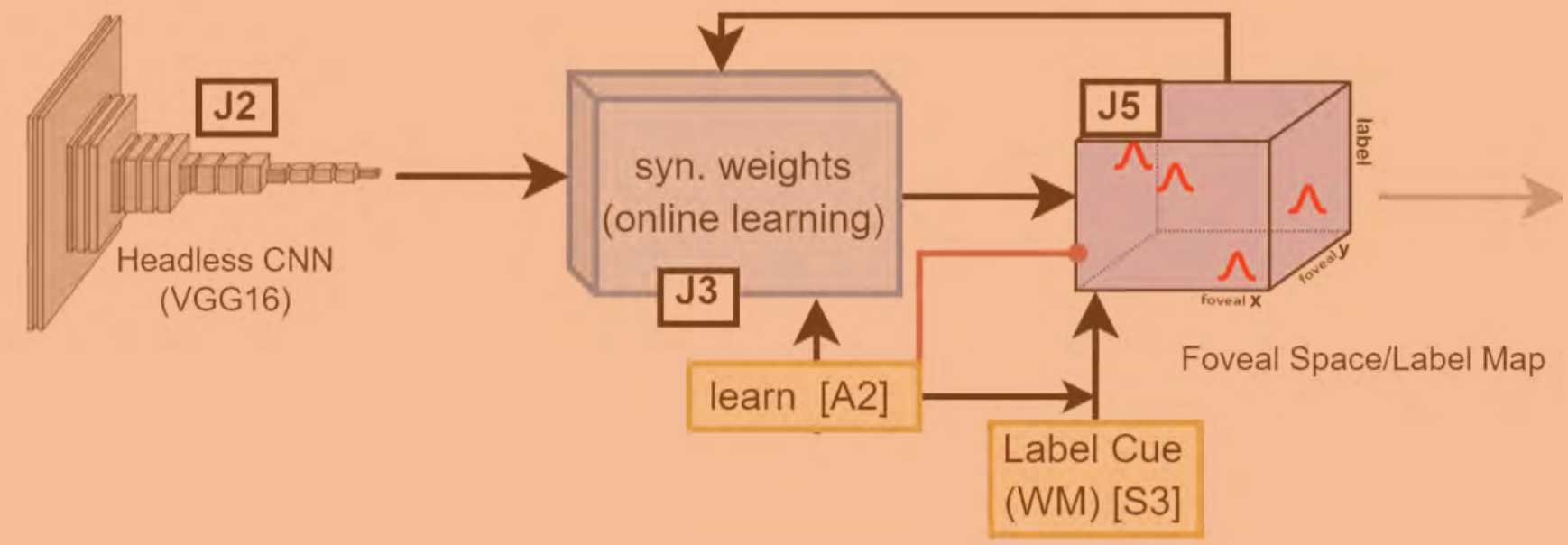


Dynamic Field Theory



Higher cognition

Interface

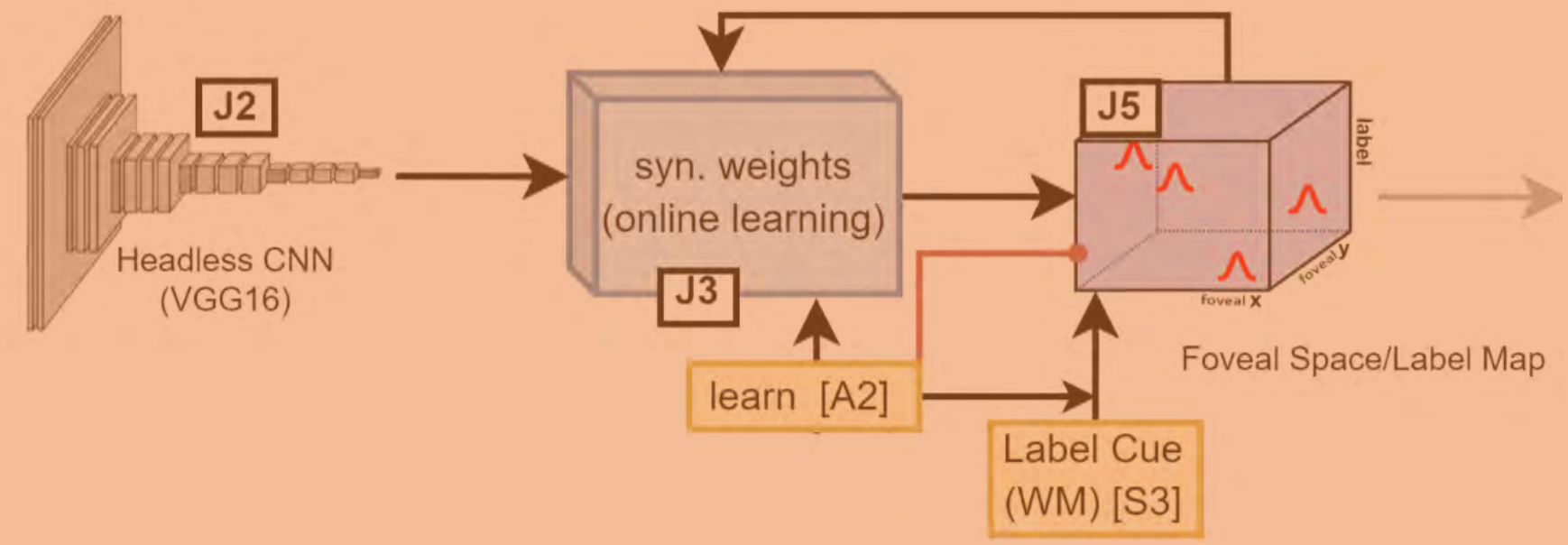


$$\tau_w \dot{w}_{m_f,u_{fsml}}(\boldsymbol{x},t) = \eta\ \sigma(u_{learn})\ y\ (y-\Theta)\ \frac{m_f(x_1,x_2,t)}{\Theta}$$

$$y = \sigma(u_{fsml}(\boldsymbol{x},t))$$

$$\tau_\Theta \dot{\Theta} = (y^2 - \Theta),$$

Interface



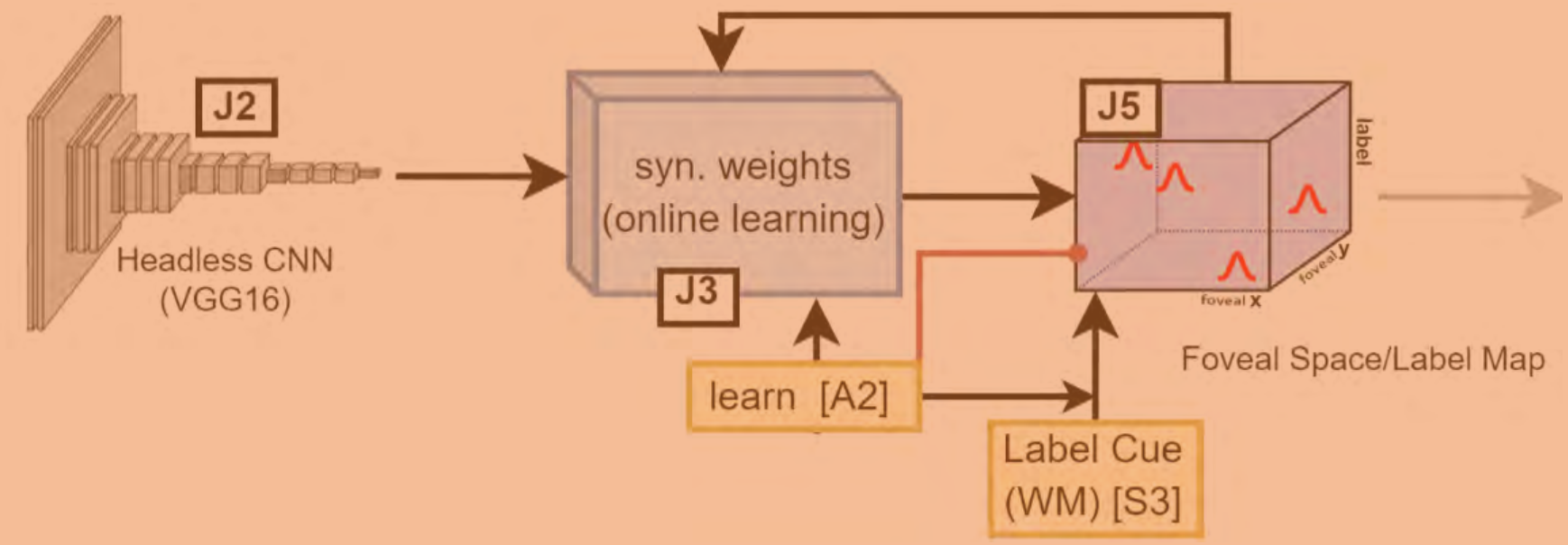
$$\tau_w \dot{w}_{m_f,u_{fsml}}(\boldsymbol{x},t) = \eta\ \sigma(u_{learn})\ y\ (y-\Theta)\ \frac{m_f(x_1,x_2,t)}{\Theta}$$

$$y = \sigma(u_{fsml}(\boldsymbol{x},t))$$

$$\tau_\Theta \dot{\Theta} = (y^2 - \Theta),$$

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity. Journal of Neuroscience

Interface



$$\tau_w \dot{w}_{m_f,u_{fsml}}(\boldsymbol{x},t) = \eta\ \sigma(u_{learn})\ y\ (y-\Theta)\ \frac{m_f(x_1,x_2,t)}{\Theta}$$

$$y = \sigma(u_{fsml}(\boldsymbol{x},t))$$

$$\tau_\Theta \dot{\Theta} = (y^2 - \Theta),$$

Lim, S., et al. (2015). Inferring learning rules from distributions of firing rates in cortical neurons. Nature neuroscience

Attention, Perception, & Psychophysics (2020) 82:775–798
https://doi.org/10.3758/s13414-019-01898-y

40 YEARS OF FEATURE INTEGRATION: SPECIAL ISSUE IN MEMORY OF ANNE TREISMAN

Scene memory and spatial inhibition in visual search

A neural dynamic process model and new experimental evidence

Raul Grieben[1] · Jan Tekülve[1] · Stephan K. U. Zibner[1] · Jonas Lins[1] · Sebastian Schneegans[2] · Gregor Schöner[1]

Published online: 11 February 2020

Abstract

Any object-oriented action requires that the object be first brought into the attentional foreground, often through visual search. Outside the laboratory, this would always take place in the presence of a scene representation acquired from ongoing visual exploration. The interaction of scene memory with visual search is still not completely understood. Feature integration theory (FIT) has shaped both research on visual search, emphasizing the scaling of search times with set size when searches entail feature conjunctions, and research on visual working memory through the change detection paradigm. Despite its neural motivation, there is no consistently neural process account of FIT in both its dimensions. We propose such an account that integrates (1) visual exploration and the building of scene memory, (2) the attentional detection of visual transients and the extraction of search cues, and (3) visual search itself. The model uses dynamic field theory in which networks of neural dynamic populations supporting stable activation states are coupled to generate sequences of processing steps. The neural architecture accounts for basic findings in visual search and proposes a concrete mechanism for the integration of working memory into the search process. In a behavioral experiment, we address the long-standing question of whether both the overall speed and the efficiency of visual search can be improved by scene memory. We find both effects and provide model fits of the behavioral results. In a second experiment, we show that the increase in efficiency is fragile, and trace that fragility to the resetting of spatial working memory.

Grieben, R., Tekülve, J., Zibner, S. K. U., Lins, J., Schneegans, S., & Schöner, G.. (2020). Scene memory and spatial inhibition in visual search. Attention, Perception, & Psychophysics.

A neural dynamic process model of combined bottom-up and top-down guidance in triple conjunction visual search

Raul Grieben (raul.grieben@ini.rub.de)
Ruhr-Universität Bochum, Institut für Neuroinformatik
Universitätsstraße 150, 44801 Bochum, Germany

Gregor Schöner (gregor.schoener@ini.rub.de)
Ruhr-Universität Bochum, Institut für Neuroinformatik
Universitätsstraße 150, 44801 Bochum, Germany

Abstract

The surprising efficiency of triple conjunction search has created a puzzle for modelers who link visual feature binding to selective attention, igniting an ongoing debate on whether features are bound with or without attention. Nordfang and Wolfe (2014) identified feature sharing and grouping as important factors in solving the puzzle and thereby established new constraints for models of visual search. Here we extend our neural dynamic model of scene perception and visual search (Grieben et al., 2020) to account for these constraints without the need for preattentive binding. By demonstrating that visual search is not only guided top-down, but that its efficiency is affected by bottom-up salience, we address a major theoretical weakness of models of conjunctive visual search (Proulx, 2007). We show how these complex interactions emerge naturally from the underlying neural dynamics.

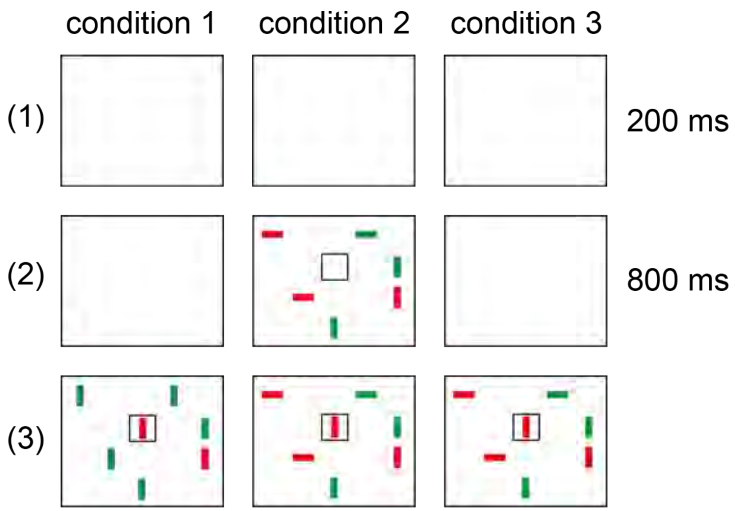
Keywords: neural dynamic process model; dynamic field theory; visual search; binding; feature sharing; grouping; triple conjunctions; bottom-up salience; top-down guidance;

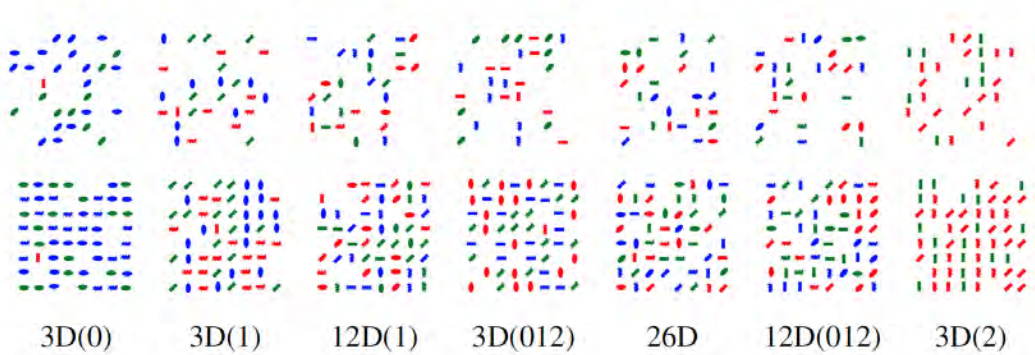
Introduction

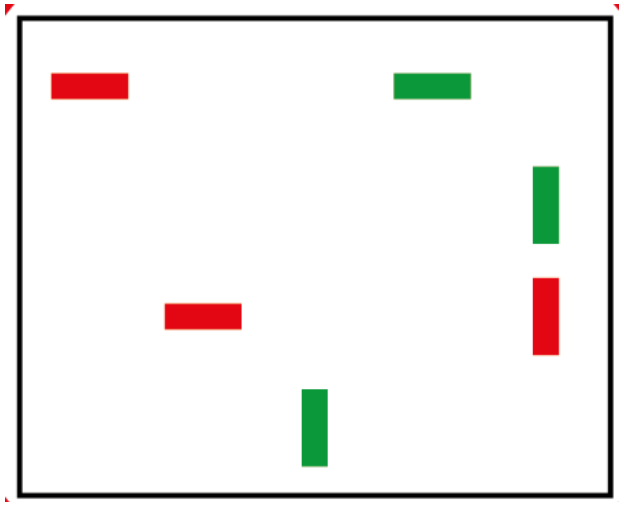
Finding an object in a natural visual scene is something we do all the time without thinking about it. Sometimes it can

Nordfang and Wolfe (2014) revisited triple conjunction searches and found evidence that both *grouping*, the number of different distractor groups in a search display, and *feature sharing*, the number of features shared between a distractor and the target, had a substantial effect on search efficiency. In the five experiments relevant for evaluation of our model (1a, 1b, 3, 4, and 6) they tested seven[1] conditions with distractor groups sharing zero features (3D(0)), one feature (3D(1), 12D(1)), two features (3D(2)) or with distractor groups composed of items with zero, one, and two shared feature values (3D(012), 12D(012), 26D). Three (3D conditions), 12 (12D conditions), or 26 (condition 26D) different distractor groups were pseudo-randomly pulled from 26 distinct triple conjunctions (*color* (red, green, blue) × *orientation* (0°, 45°, 90°) × *shape* (rectangular, oval, jagged)). The distribution of features in each condition was constrained: In condition 3D(0), no distractor could share any feature with the target. In conditions 3D(1), 3D(012), 12D(1), 12D(012), and 26D, each feature type was present in 1/3 of the items (1/3 were red, 1/3 green, and 1/3 blue, and so on). In condition 3D(2), the target

Grieben, R., & Schöner, G.. (2021). A neural dynamic process model of combined bottom-up and top-down guidance in triple conjunction visual search. Proceedings of the 43rd Annual Conference of the Cognitive Science Society.

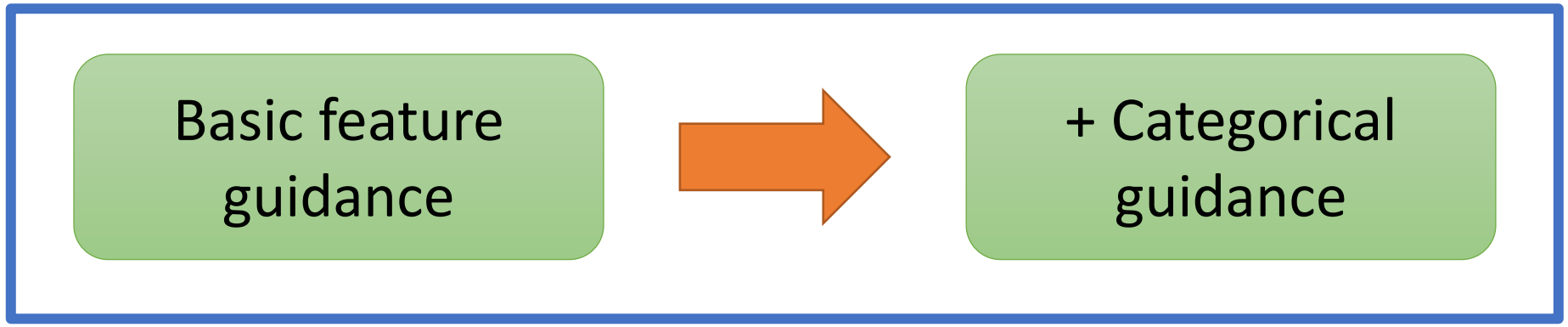

condition 1
condition 2
condition 3
(1)
(2)
(3)
200 ms
800 ms

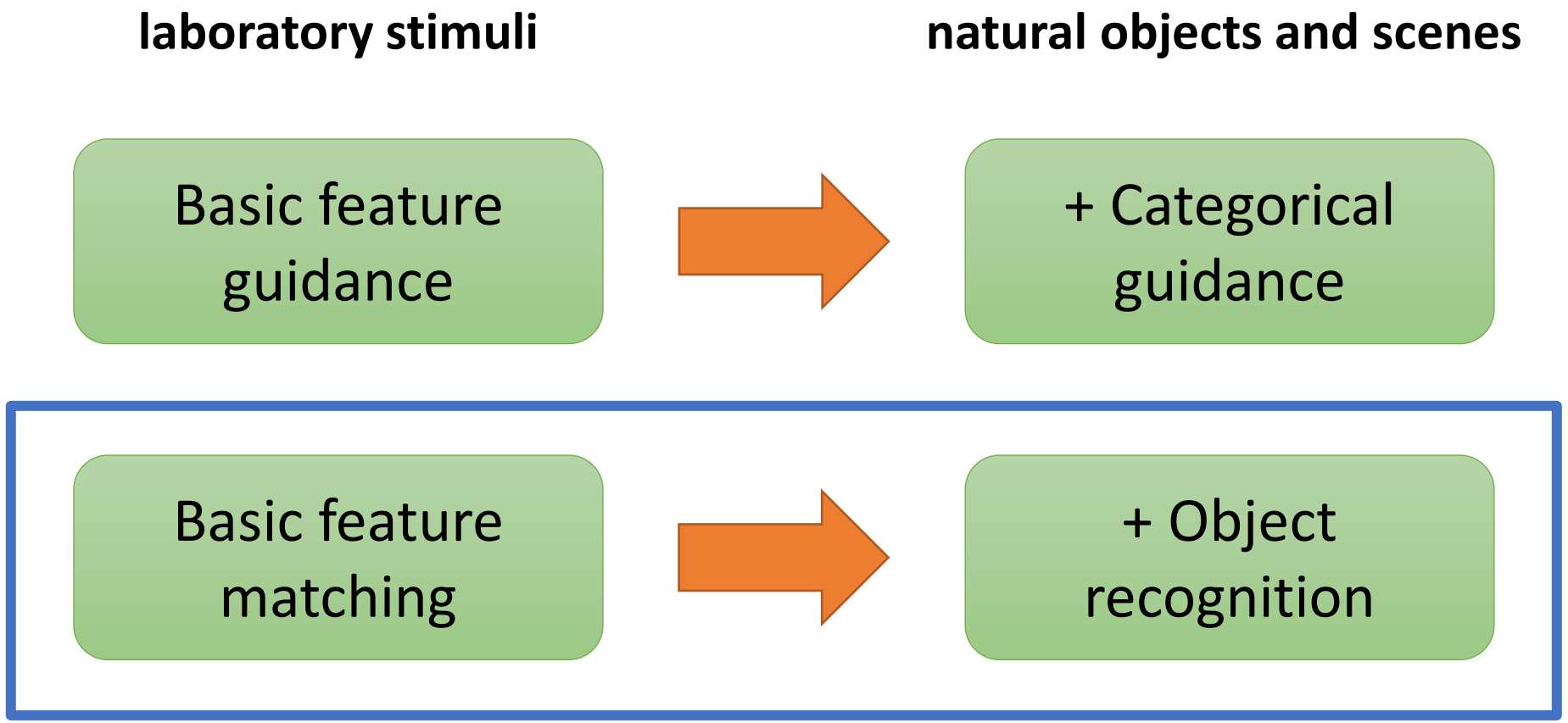

3D(0)
3D(1)
12D(1)
3D(012)
26D
12D(012)
3D(2)



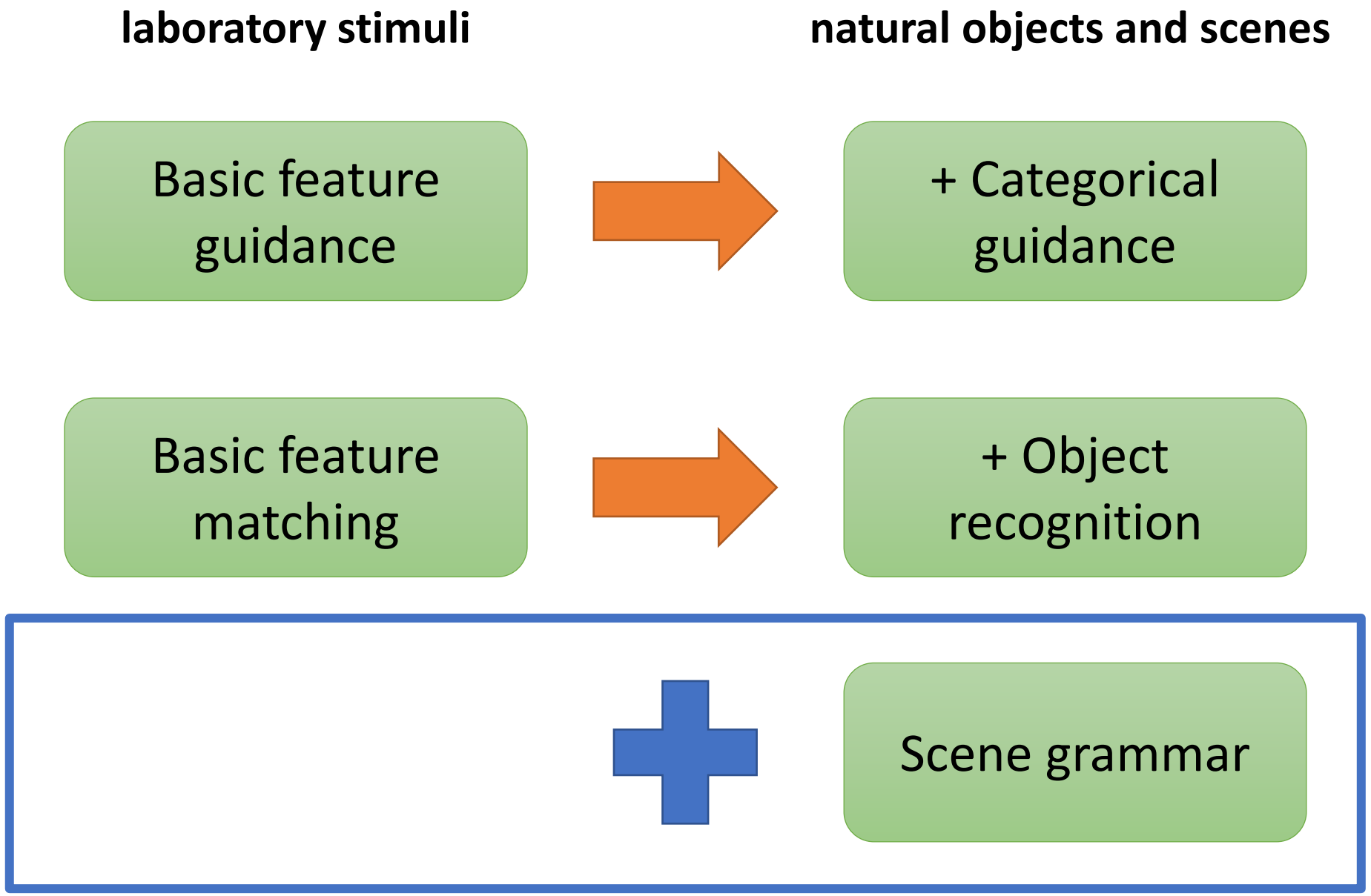

PRIVAT

laboratory stimuli **natural objects and scenes**

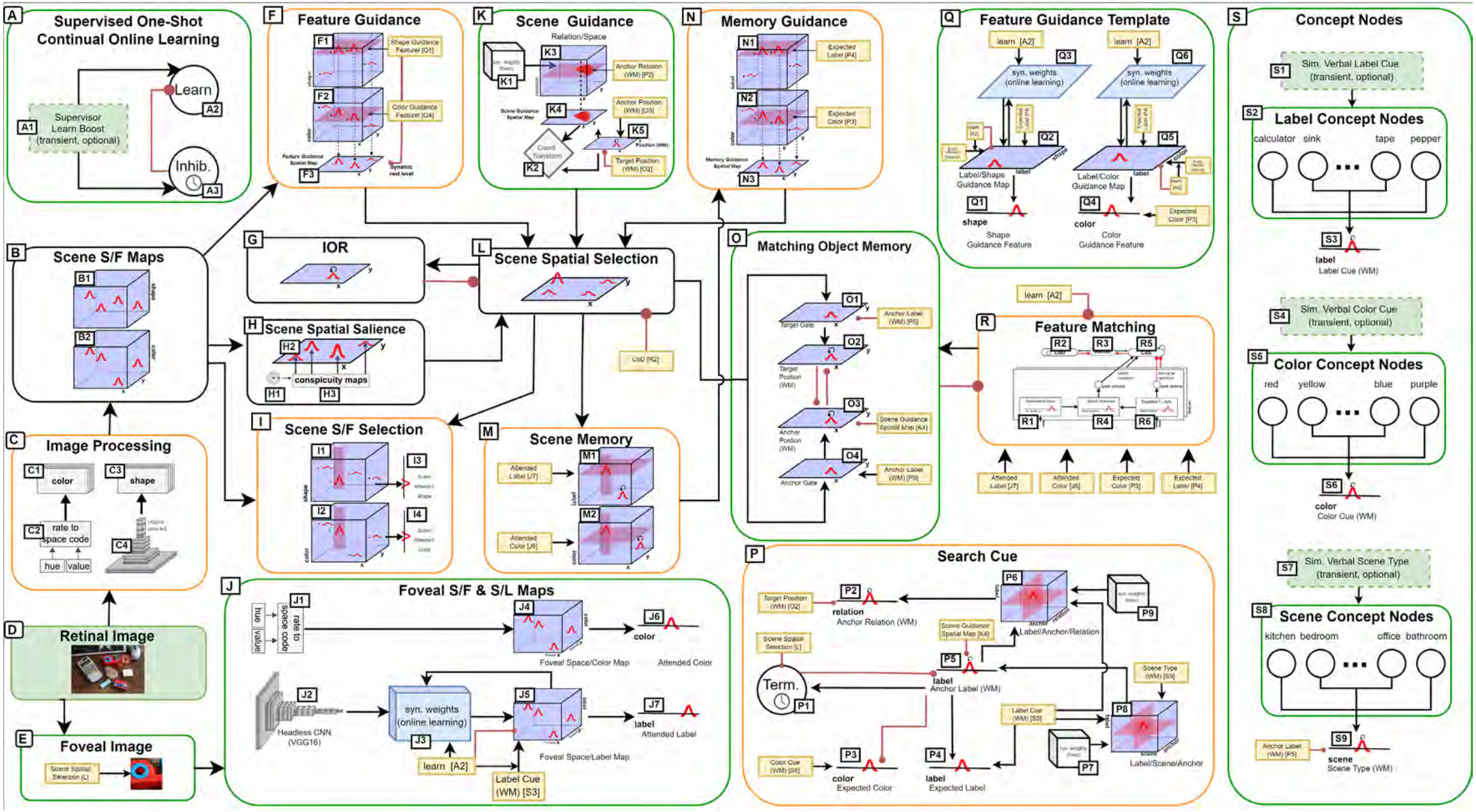



laboratory stimuli
natural objects and scenes
Basic feature guidance
+ Categorical guidance
Basic feature matching
+ Object recognition

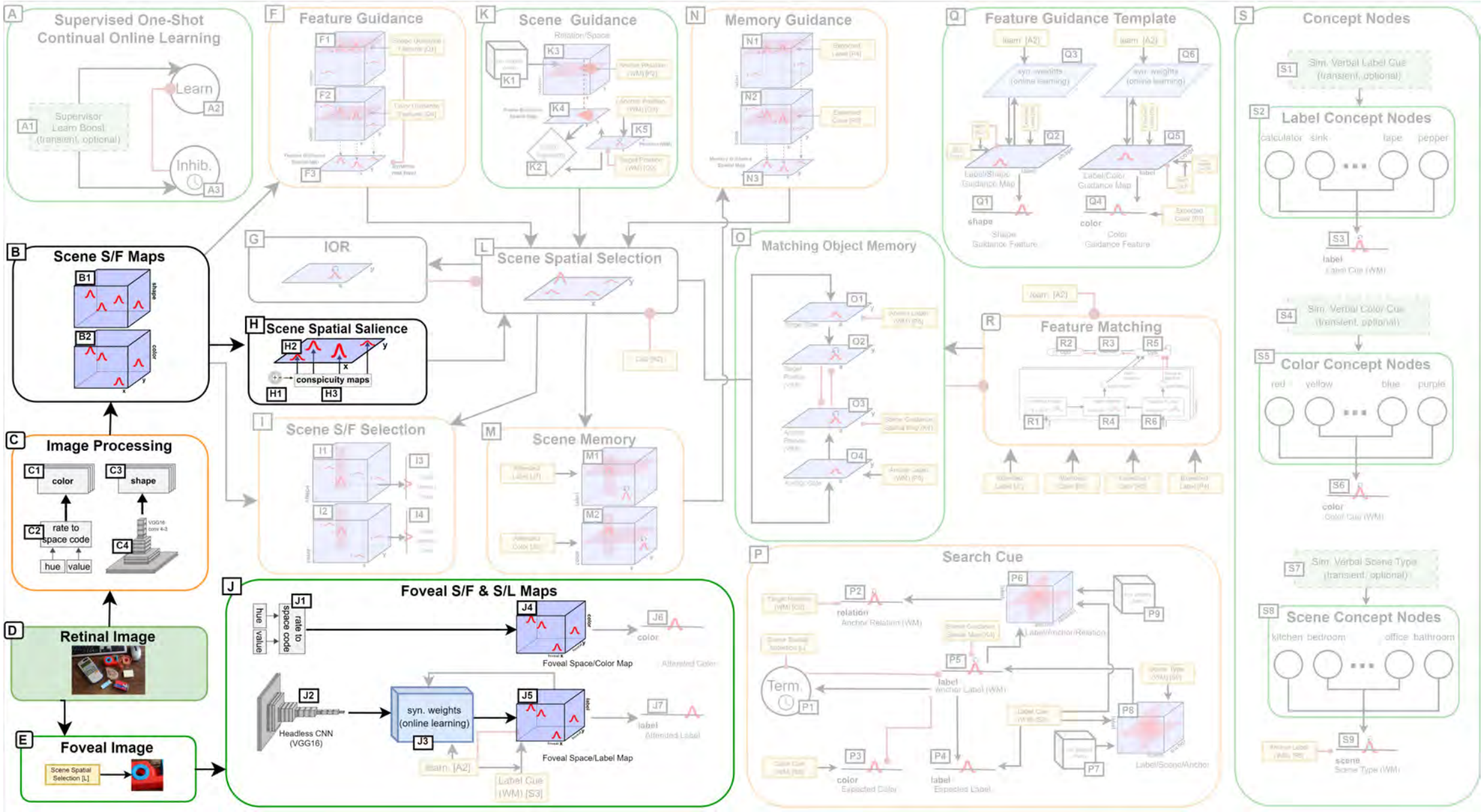
Boettcher, S. E., Draschkow, D., Dienhart, E., & Vo, M. L.- H. (2018). Anchoring visual search in scenes. Journal of vision


laboratory stimuli
natural objects and scenes
Basic feature guidance
+ Categorical guidance
Basic feature matching
+ Object recognition
Scene grammar

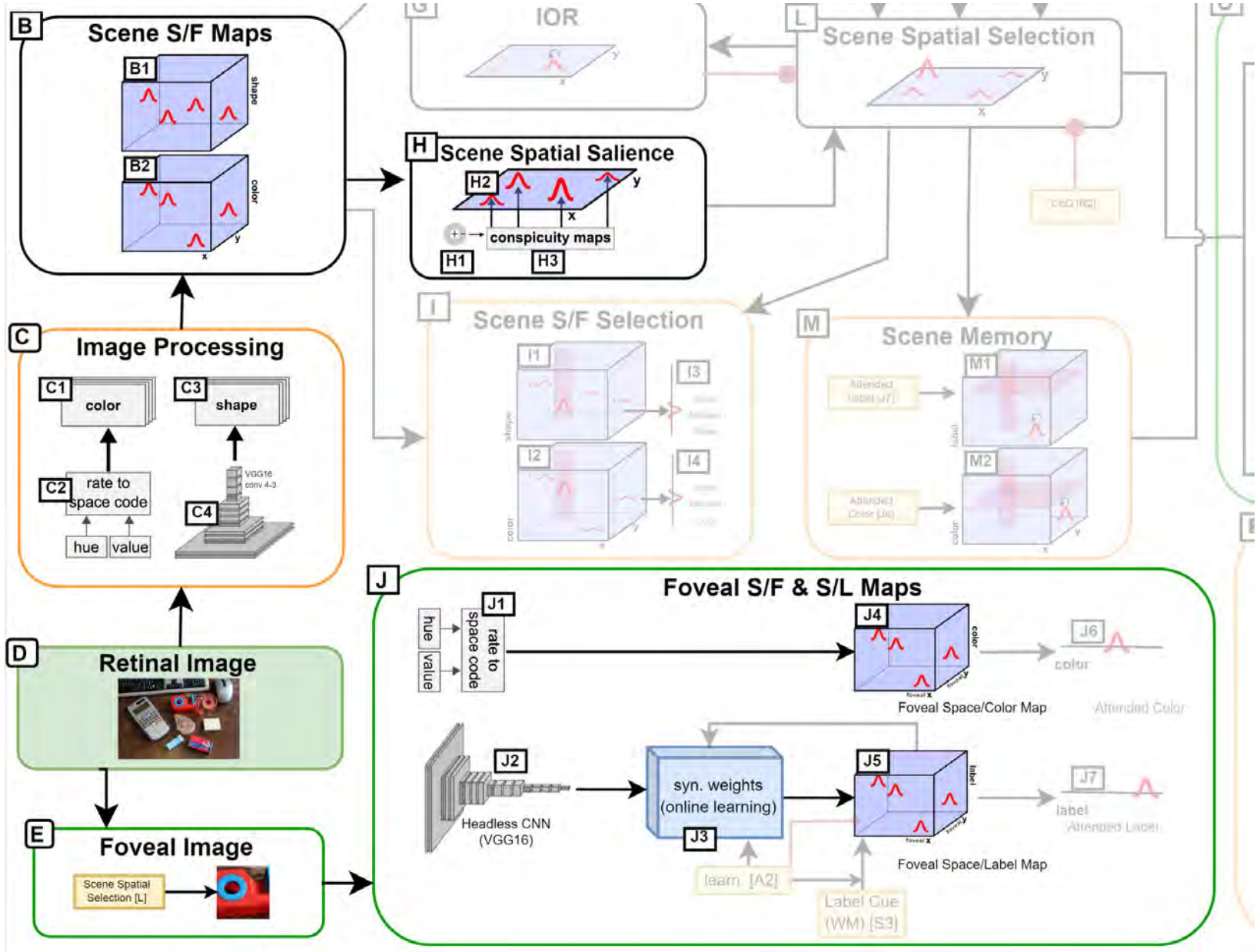
The neural process model



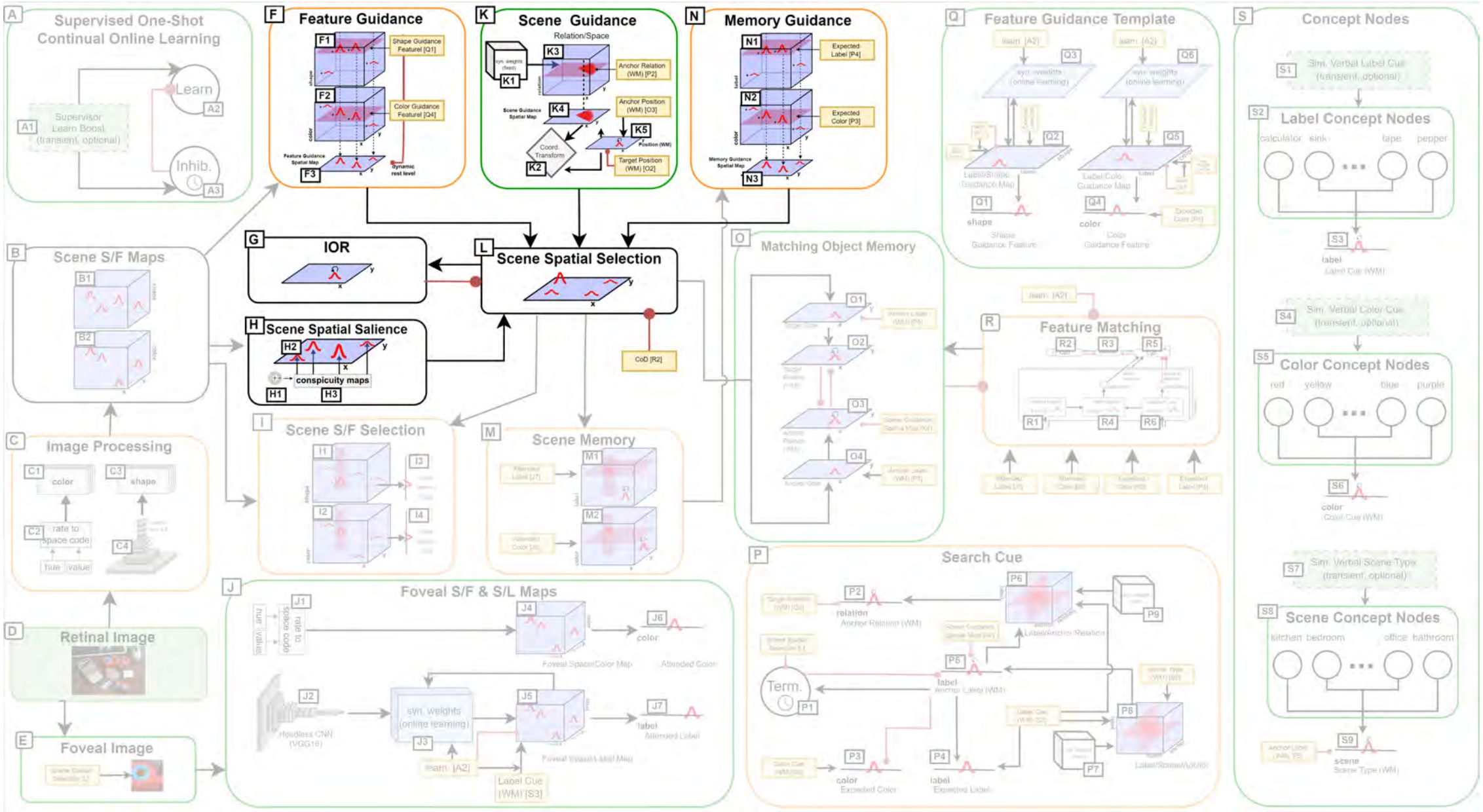
Feed-forward feature and salience maps



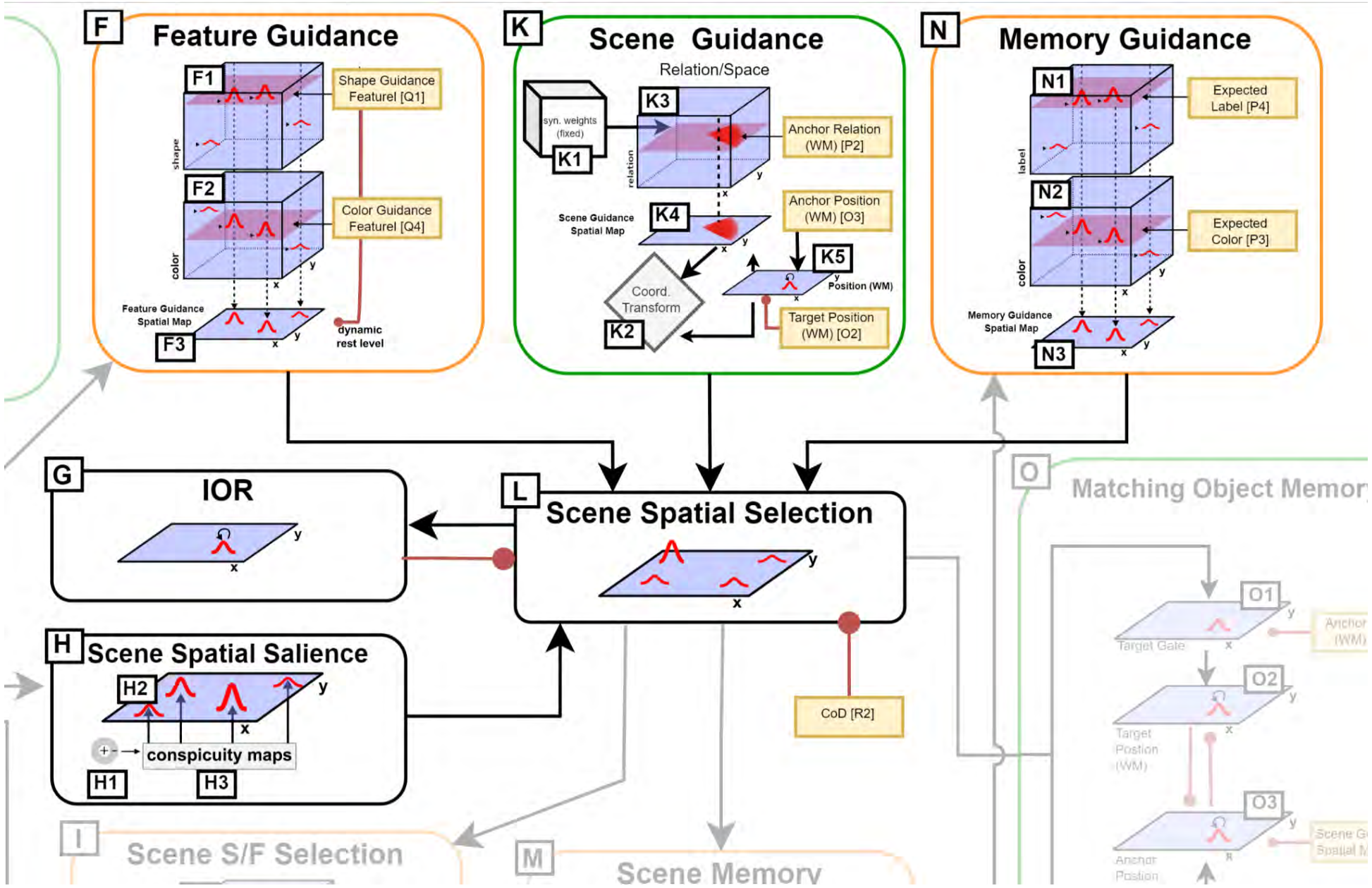
Feed-forward feature and salience maps



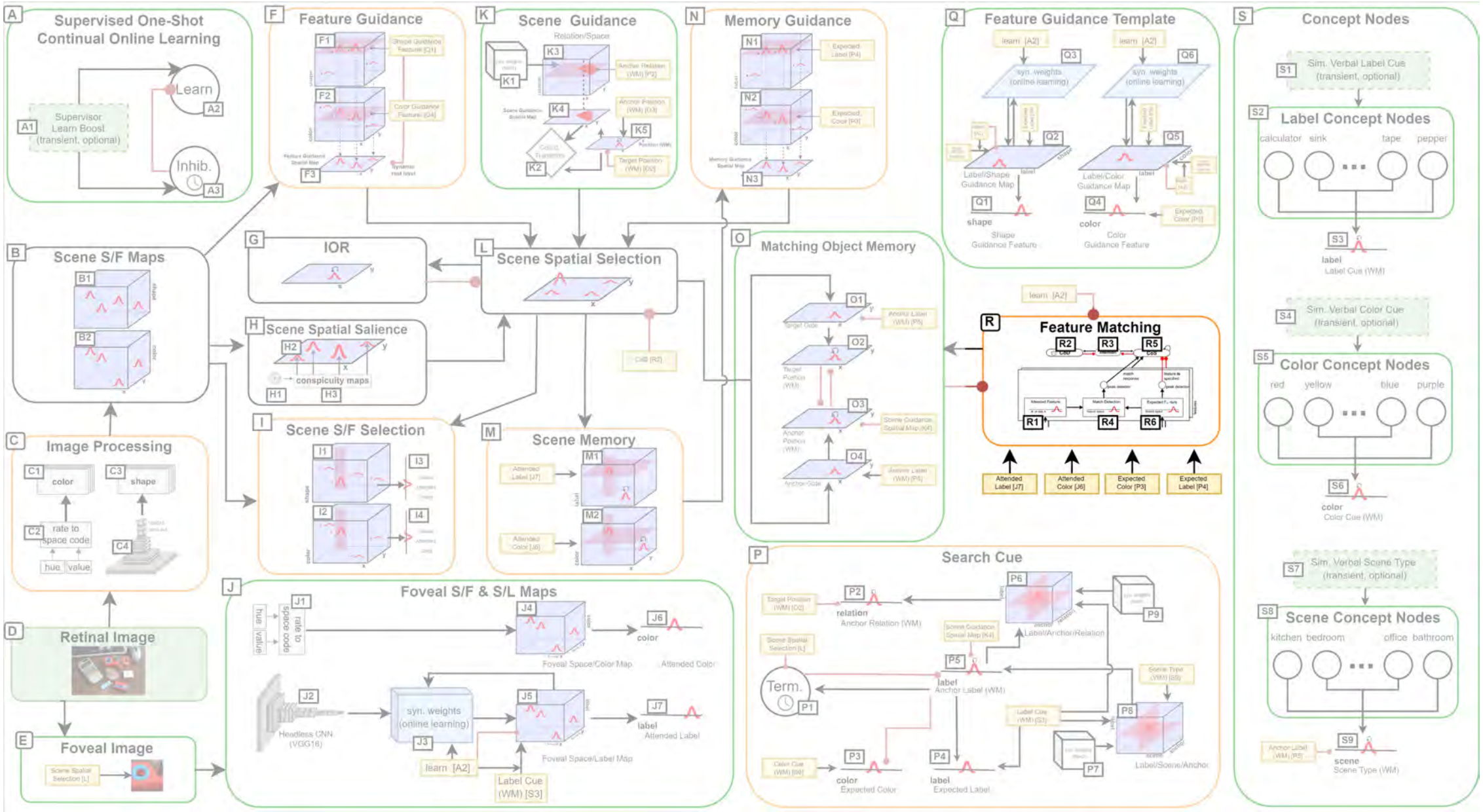
Attentional selection through biased competition



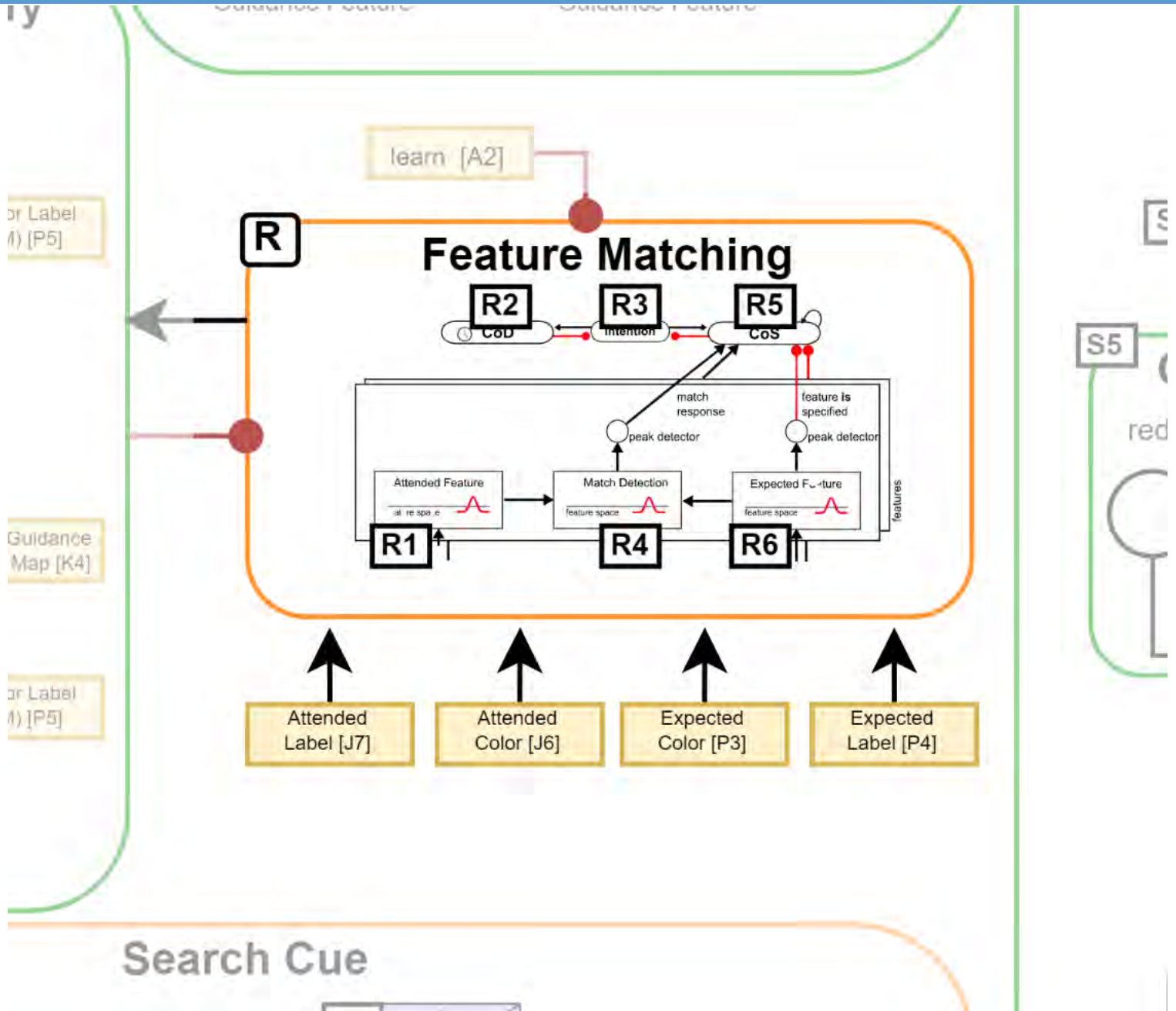
Attentional selection through biased competition



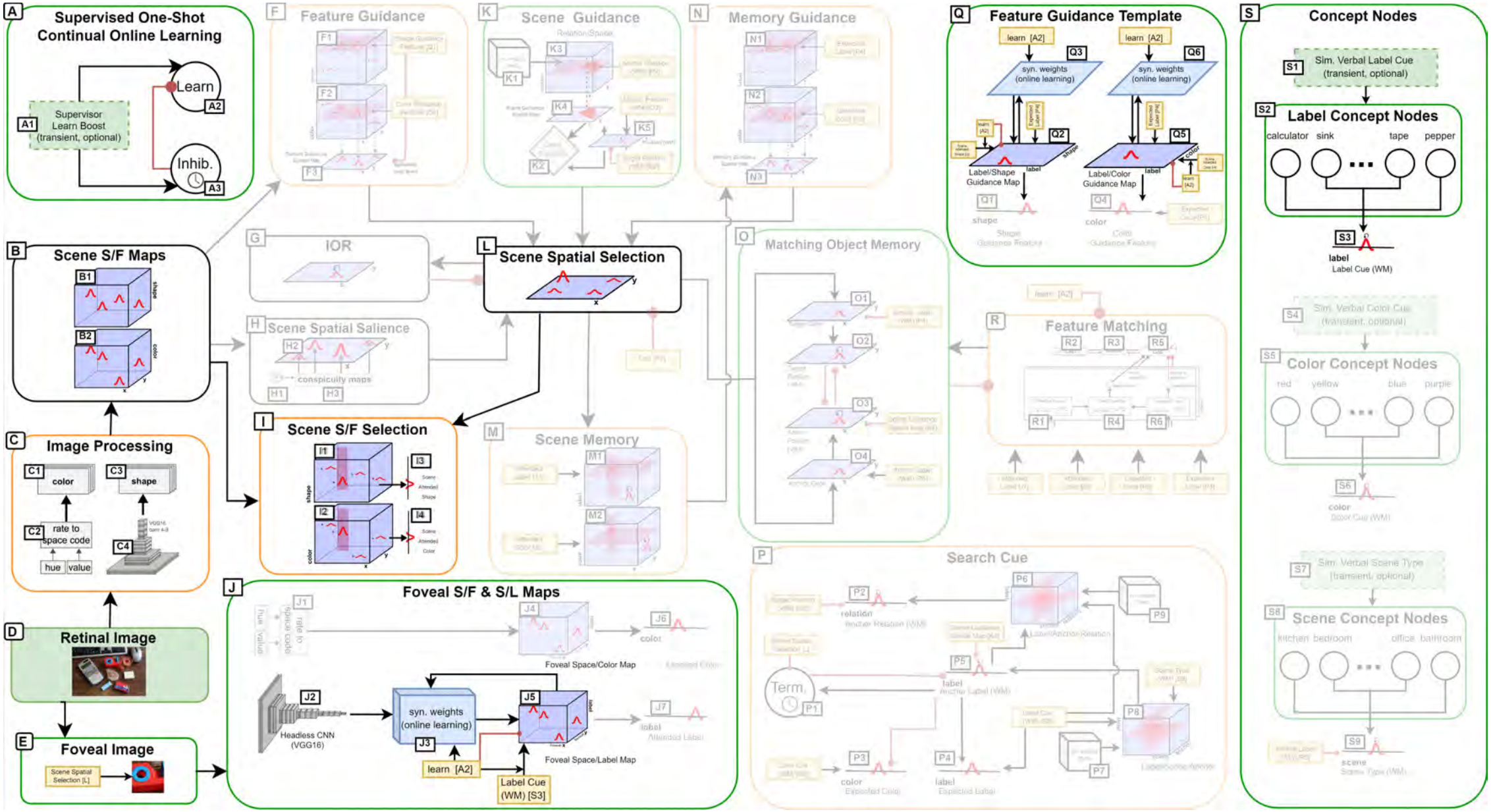
Match detection



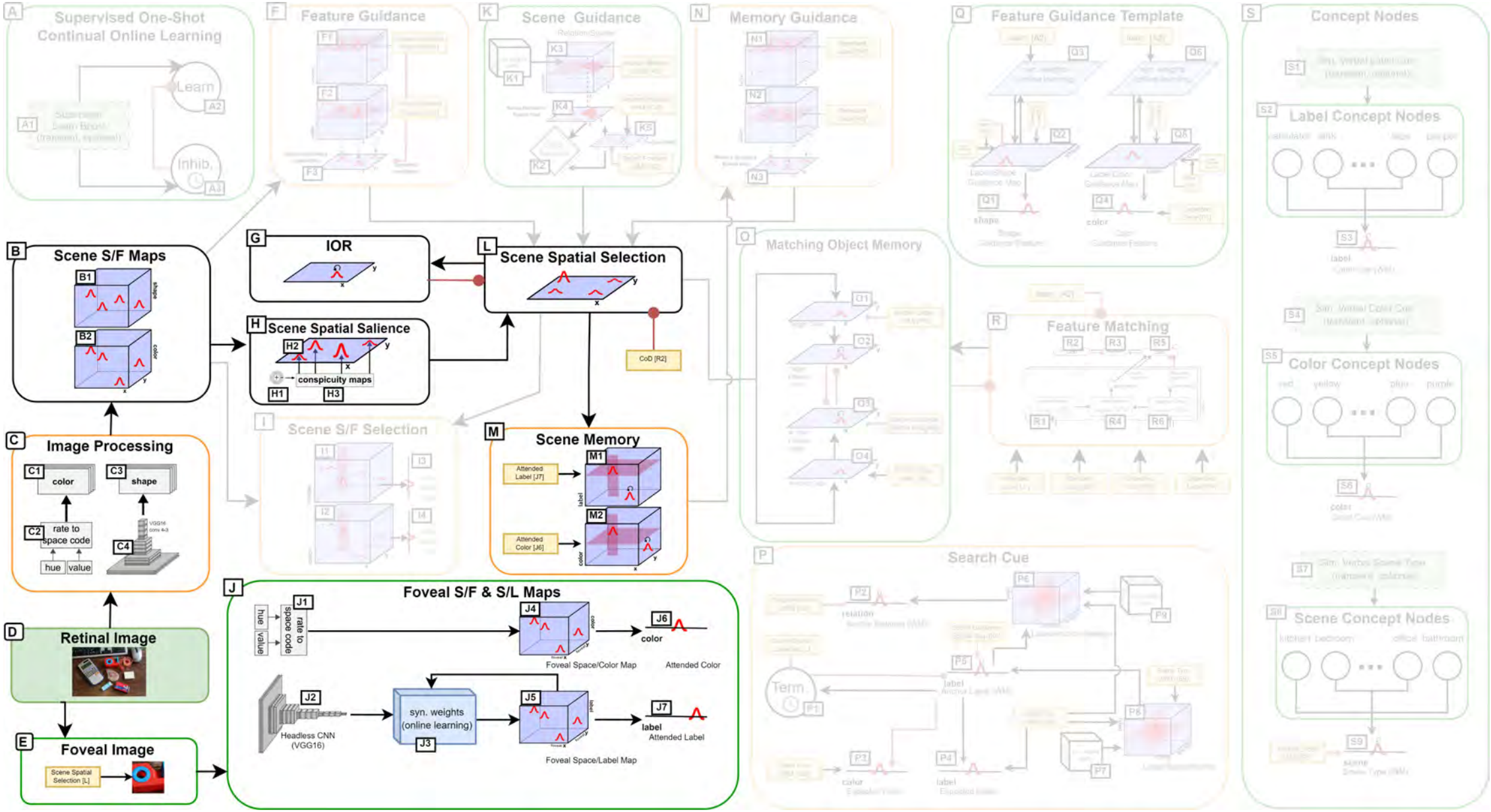
Match detection



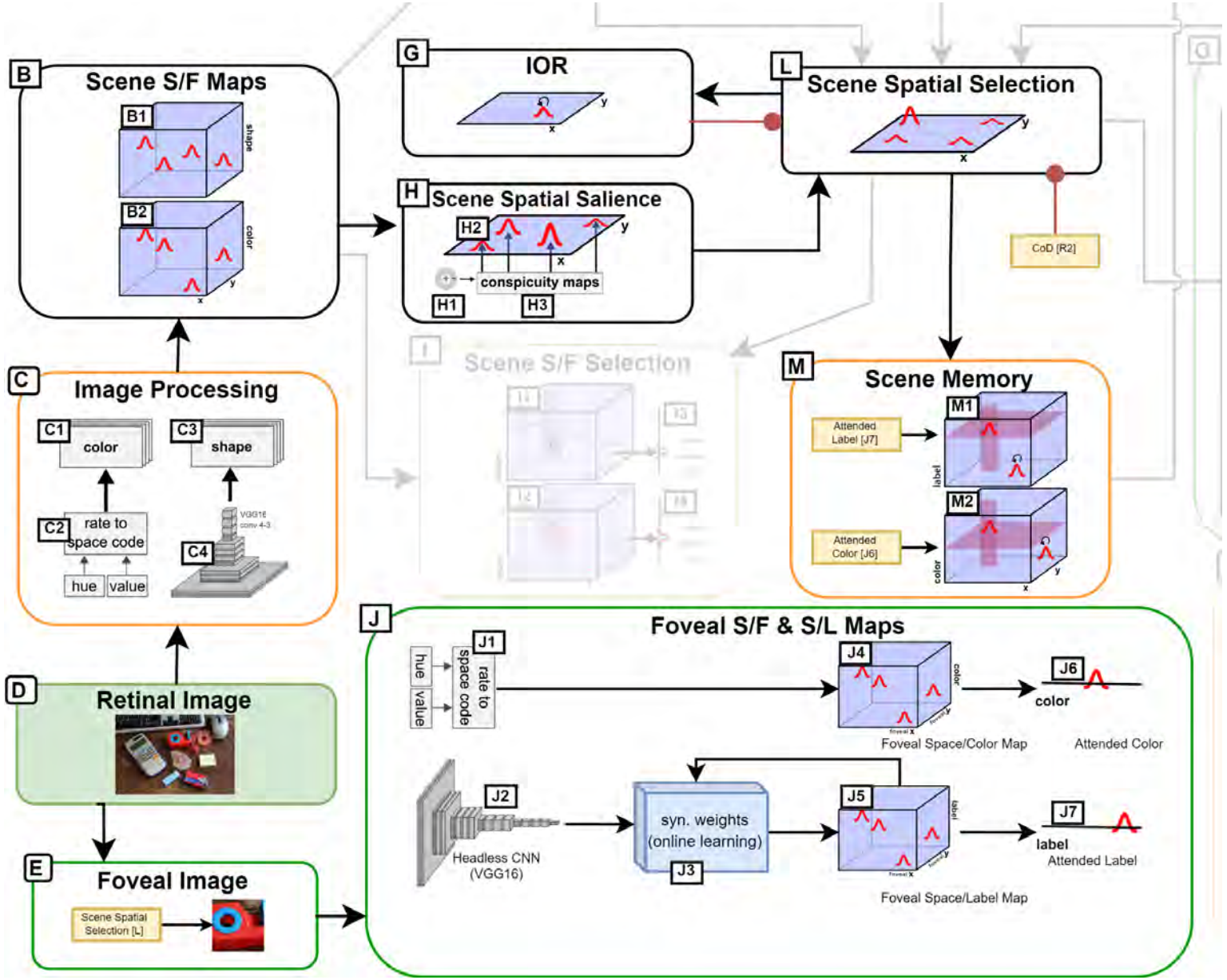
Learning


A
Supervised One-Shot Continual Online Learning
Learn
A2
Supervisor Learn Boost (transient, optional)
A1
Inhib.
A3
F
Feature Guidance
K
Scene Guidance
N
Memory Guidance
Q
Feature Guidance Template
learn [A2]
Q3
syn. weights (online learning)
Q6
Q2
Q5
Label/Shape Guidance Map
Label/Color Guidance Map
Q1
shape
Q4
color
S
Concept Nodes
S1
Sim. Verbal Label Cue (transient, optional)
S2
Label Concept Nodes
calculator
sink
tape
pepper
S3
label
Label Cue (WM)
B
Scene S/F Maps
B1
shape
B2
color
G
IOR
L
Scene Spatial Selection
O
Matching Object Memory
O1
O2
O3
O4
H
Scene Spatial Salience
conspicuity maps
H1
H3
H2
R
Feature Matching
R1
R2
R3
R4
R5
R6
S4
Sim. Verbal Color Cue (transient, optional)
S5
Color Concept Nodes
red
yellow
blue
purple
S6
color
I
Scene S/F Selection
I1
I2
I3
I4
M
Scene Memory
M1
M2
C
Image Processing
C1
color
C3
shape
C2
rate to space code
hue
value
C4
D
Retinal Image
E
Foveal Image
Scene Spatial Selection [L]
J
Foveal S/F & S/L Maps
J1
rate to space code
J4
J6
color
Foveal Space/Color Map
J2
Headless CNN (VGG16)
syn. weights (online learning)
J3
J5
J7
label
Foveal Space/Label Map
learn [A2]
Label Cue (WM) [S3]
P
Search Cue
P2
relation
P6
P9
P5
label
Term.
P1
P8
P3
color
P4
label
P7
S7
Sim. Verbal Scene Type (transient, optional)
S8
Scene Concept Nodes
kitchen
bedroom
office
bathroom
S9
scene

Autonomous visual exploration

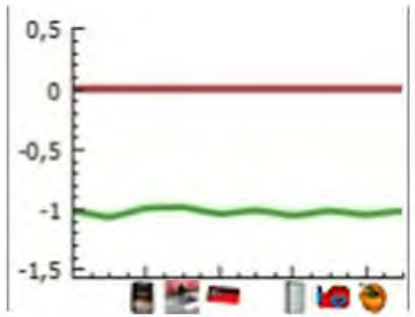


Autonomous visual exploration

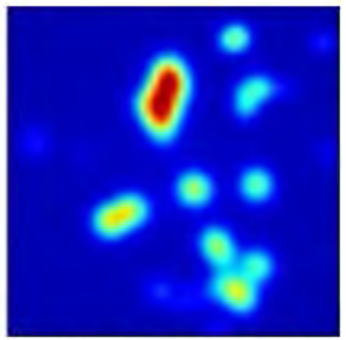


Camera Image

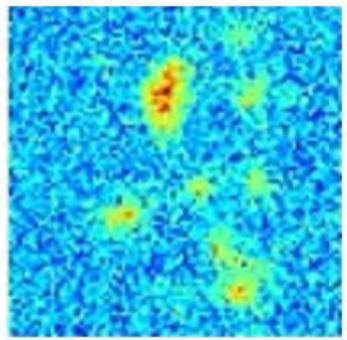
Foveal Image



Attended Label

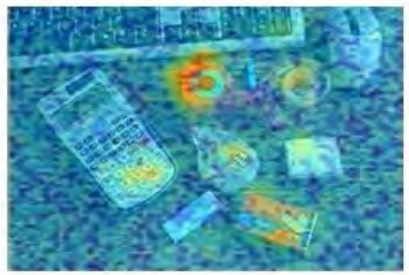


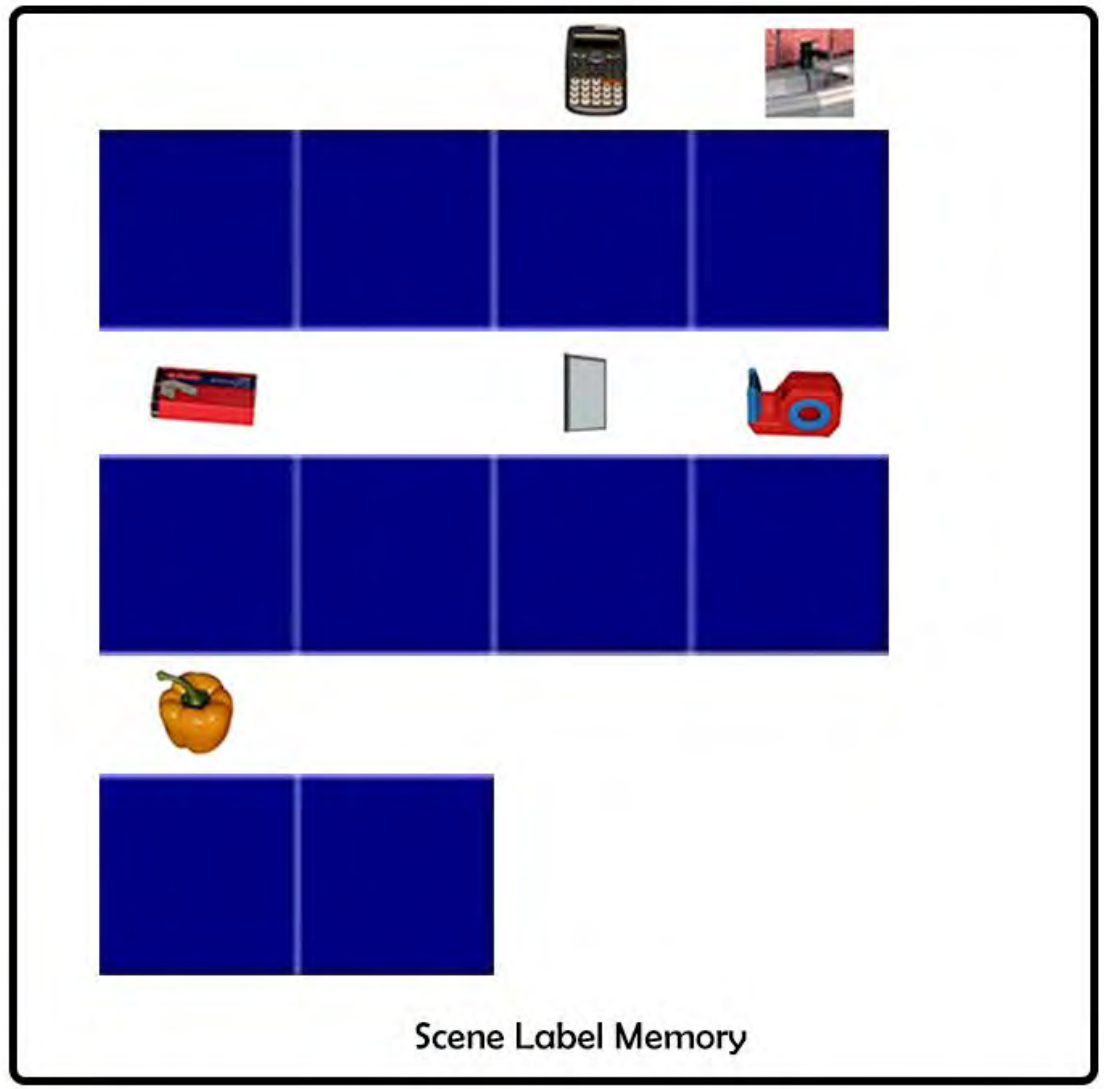
Attention (Input)



Attention (Activation)

Attention (Sig. Activation)

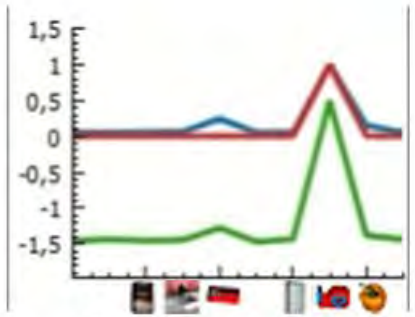




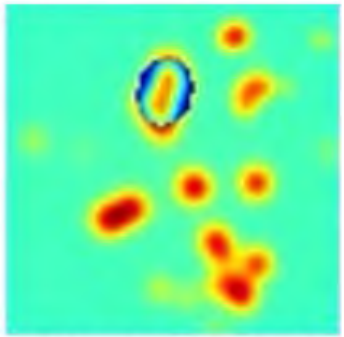
Scene Label Memory

Camera Image

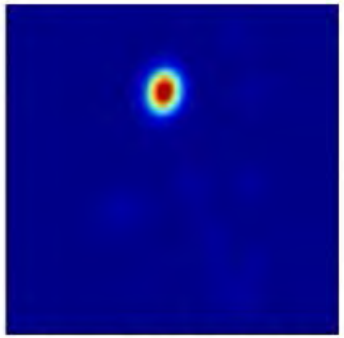
Foveal Image



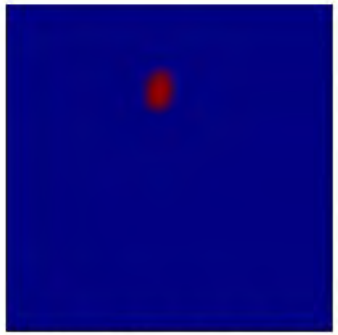
Attended Label



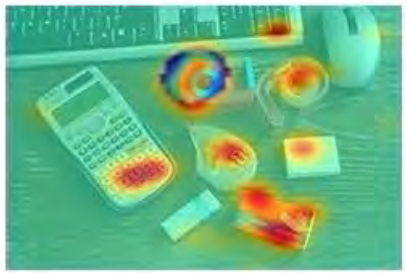
Attention (Input)

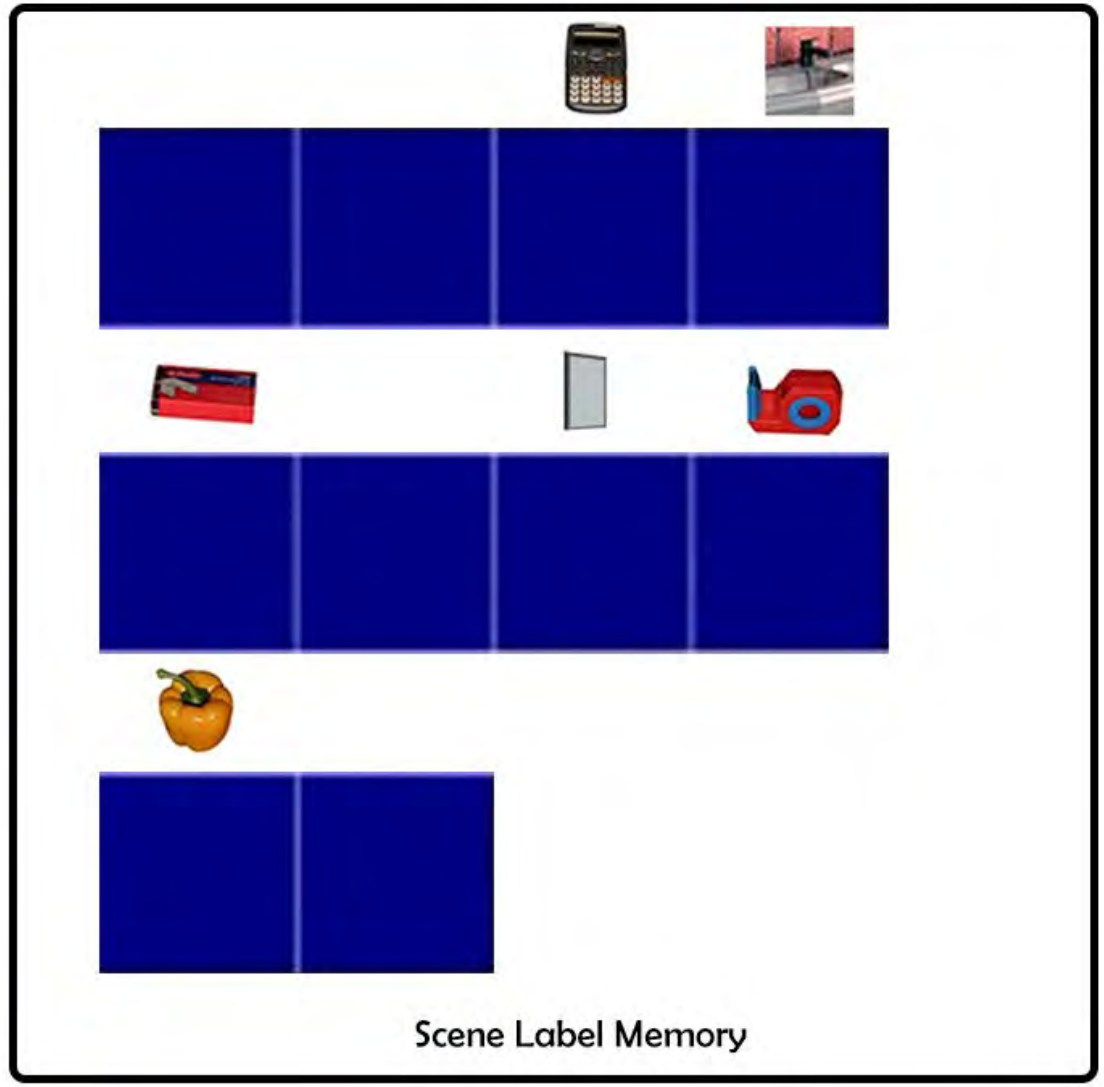


Attention (Activation)



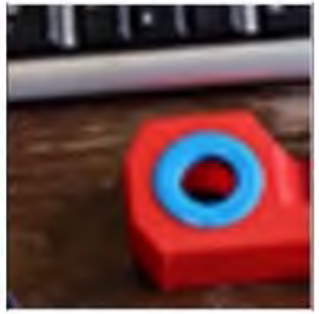
Attention (Sig. Activation)



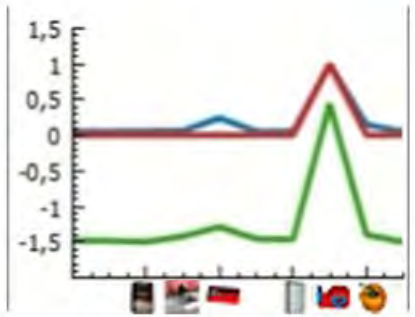


Scene Label Memory

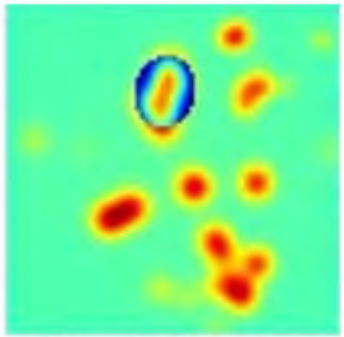
Camera Image



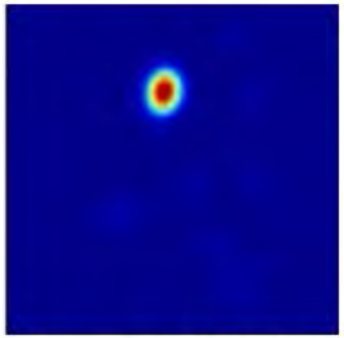
Foveal Image



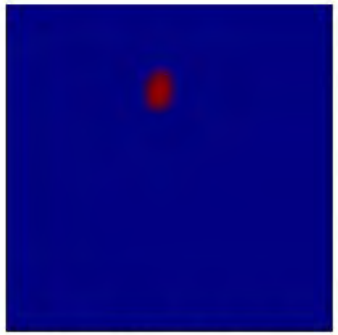
Attended Label



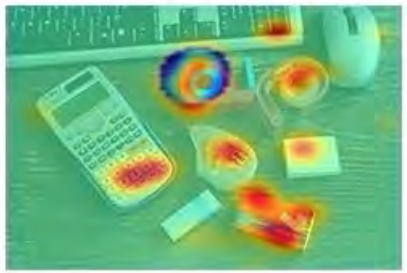
Attention (Input)

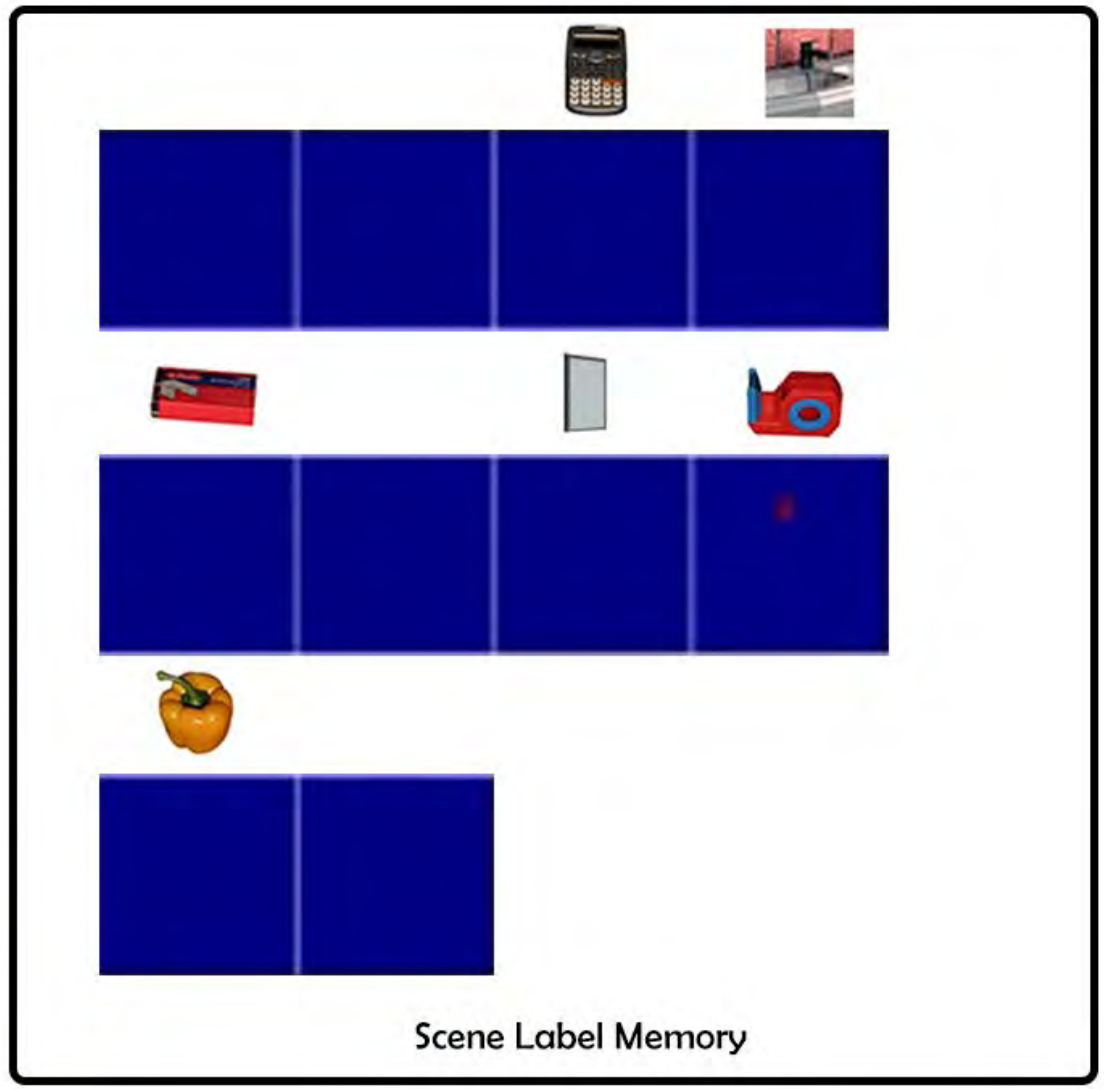


Attention (Activation)



Attention (Sig. Activation)

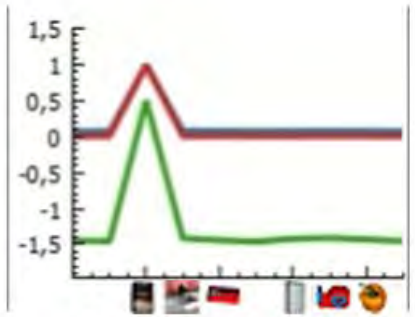




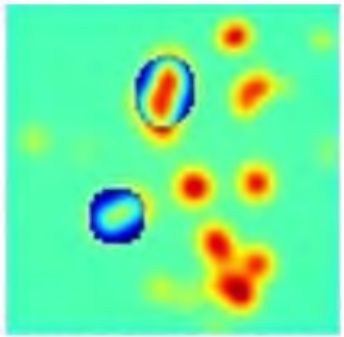
Scene Label Memory

Camera Image

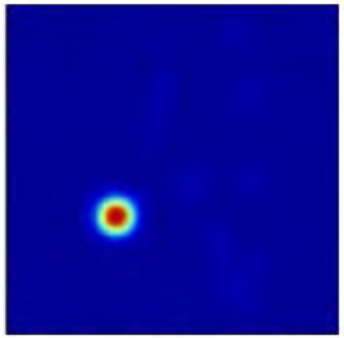
Foveal Image



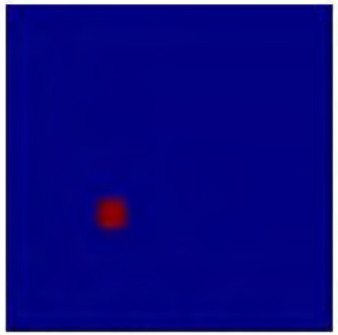
Attended Label



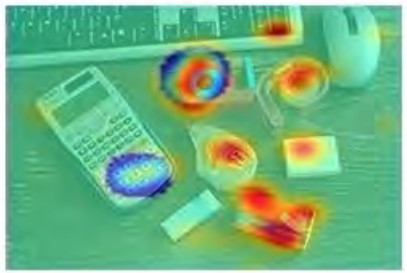
Attention (Input)

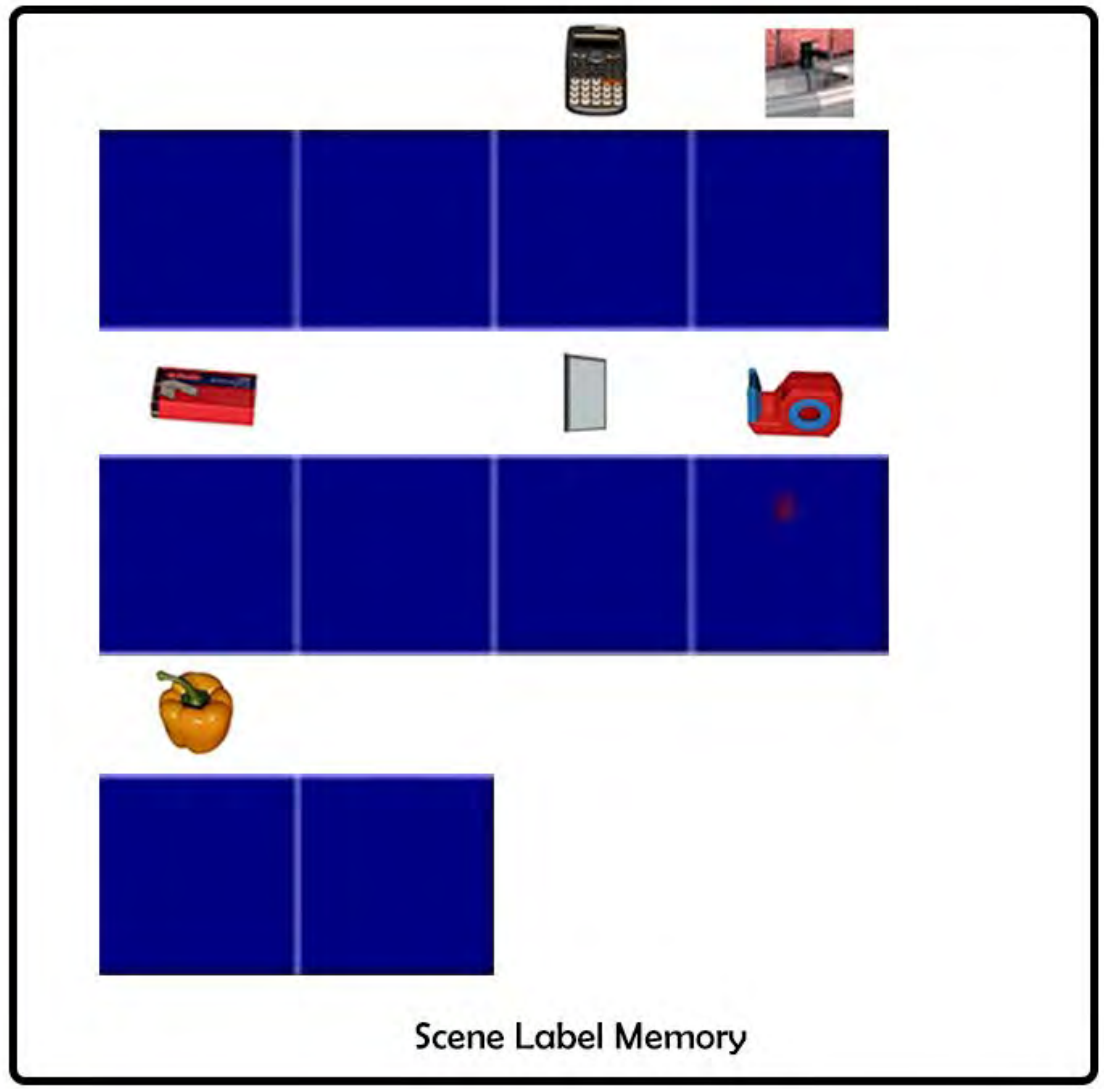


Attention (Activation)



Attention (Sig. Activation)

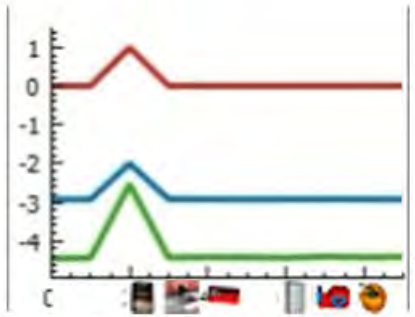




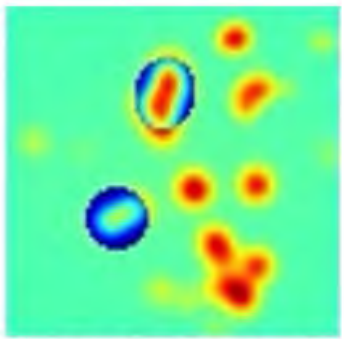
Scene Label Memory

Camera Image

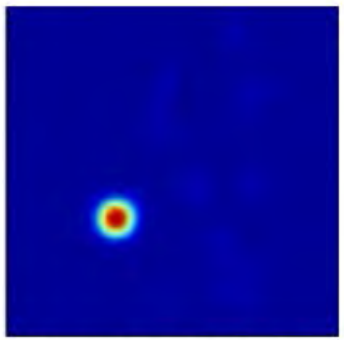
Foveal Image



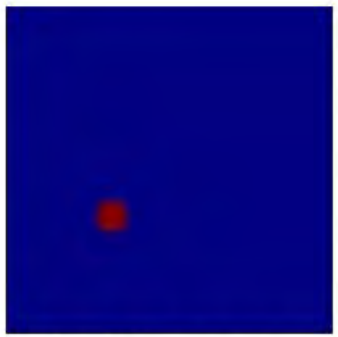
Attended Label



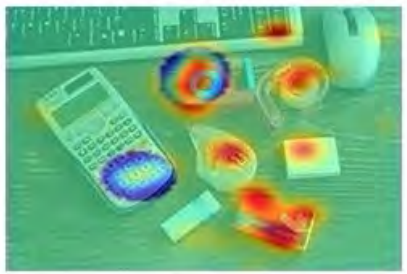
Attention (Input)

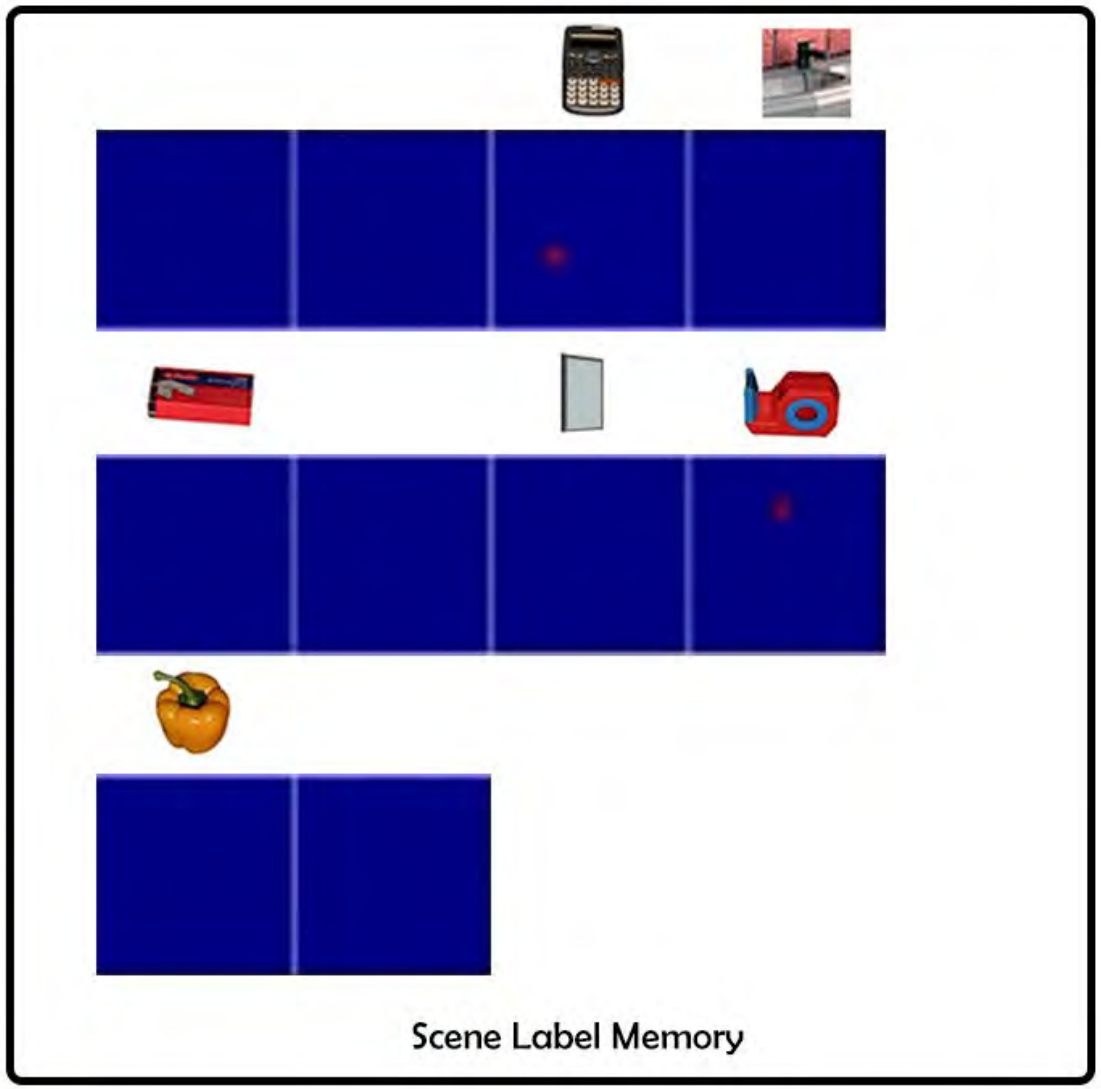


Attention (Activation)



Attention (Sig. Activation)

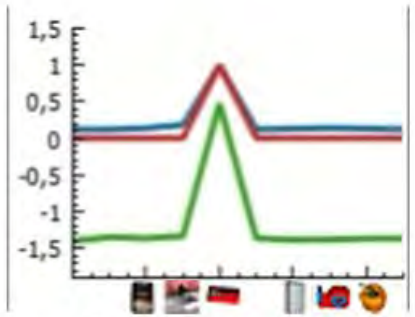




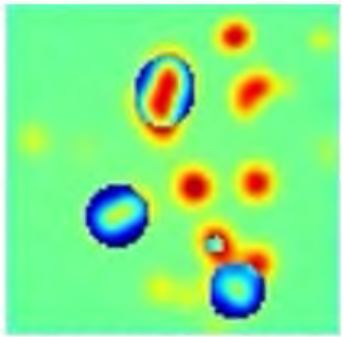
Scene Label Memory

Camera Image

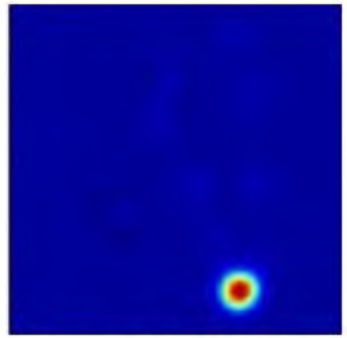
Foveal Image



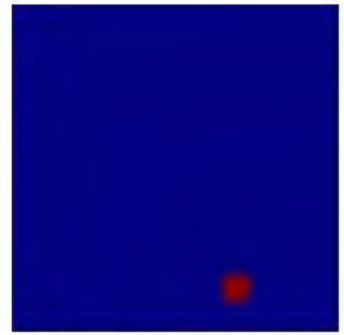
Attended Label



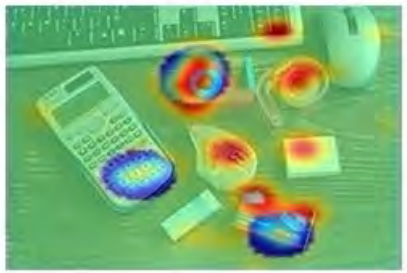
Attention (Input)

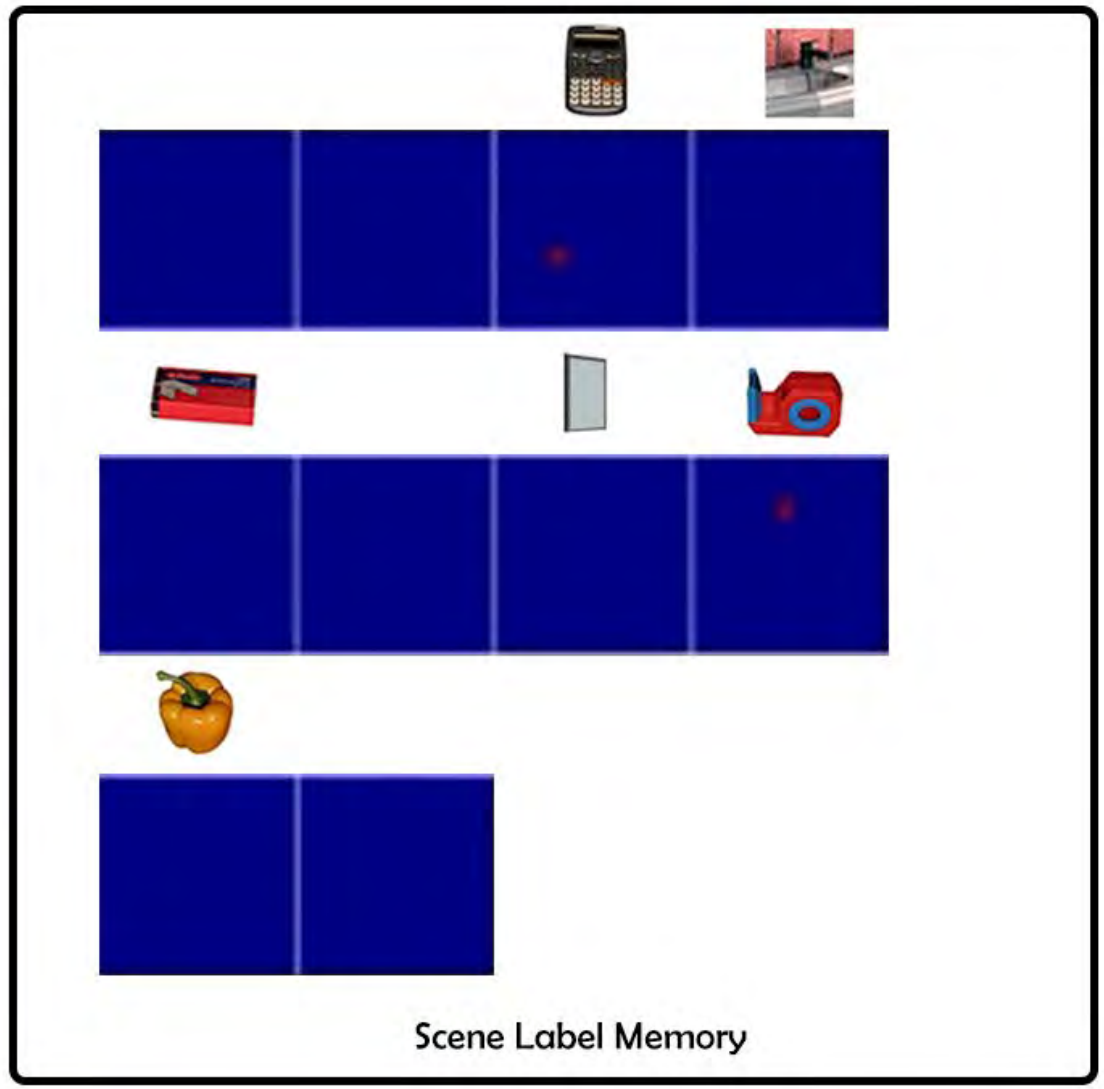


Attention (Activation)



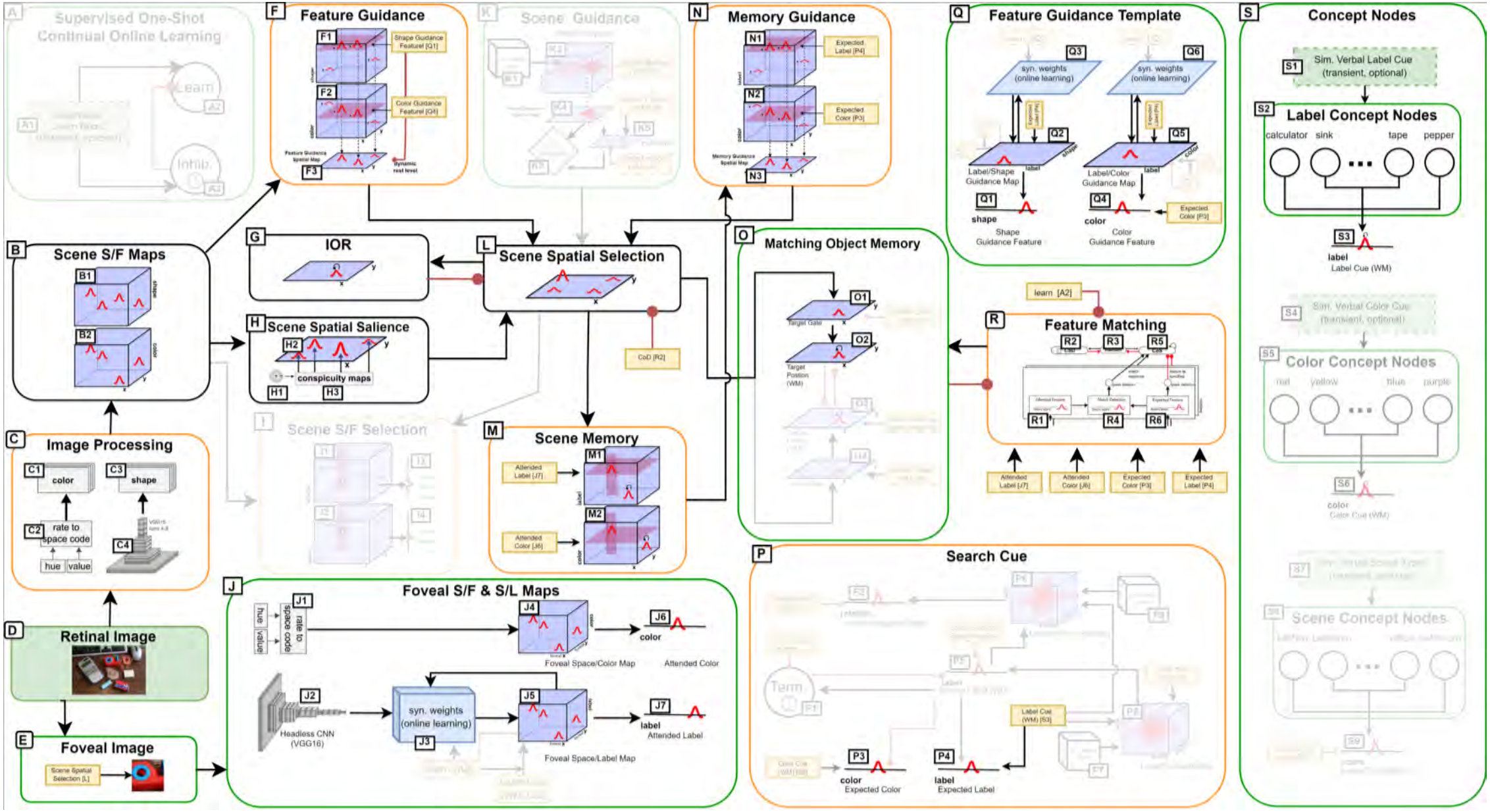
Attention (Sig. Activation)





Scene Label Memory

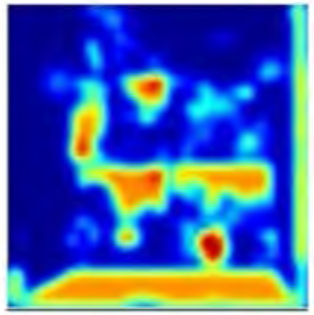
Categorical visual search



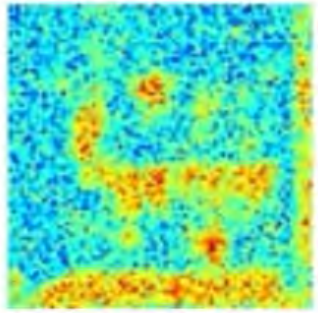
Camera Image

Foveal Image

Target Position (WM)

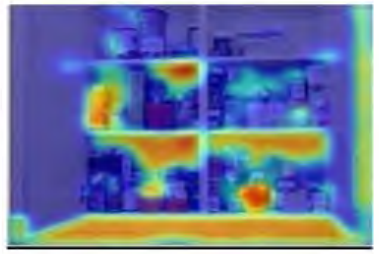


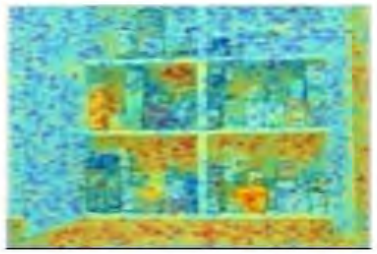
Attention (Input)

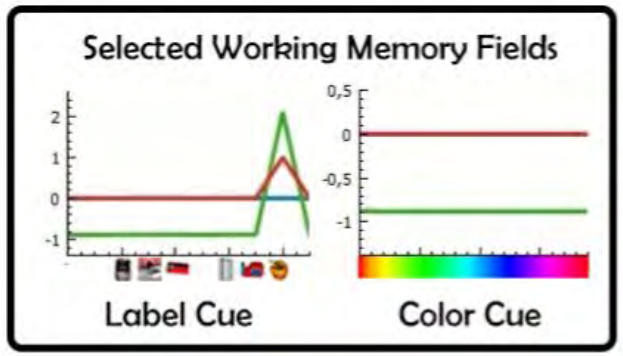


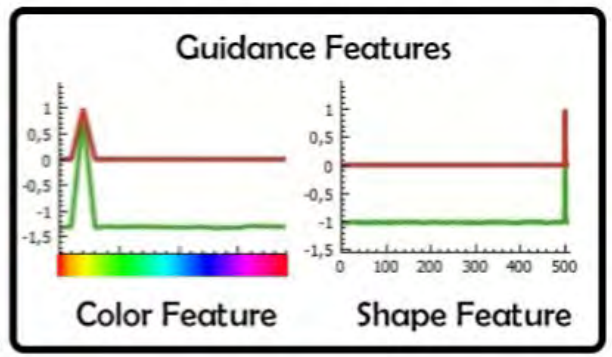
Attention (Activation)

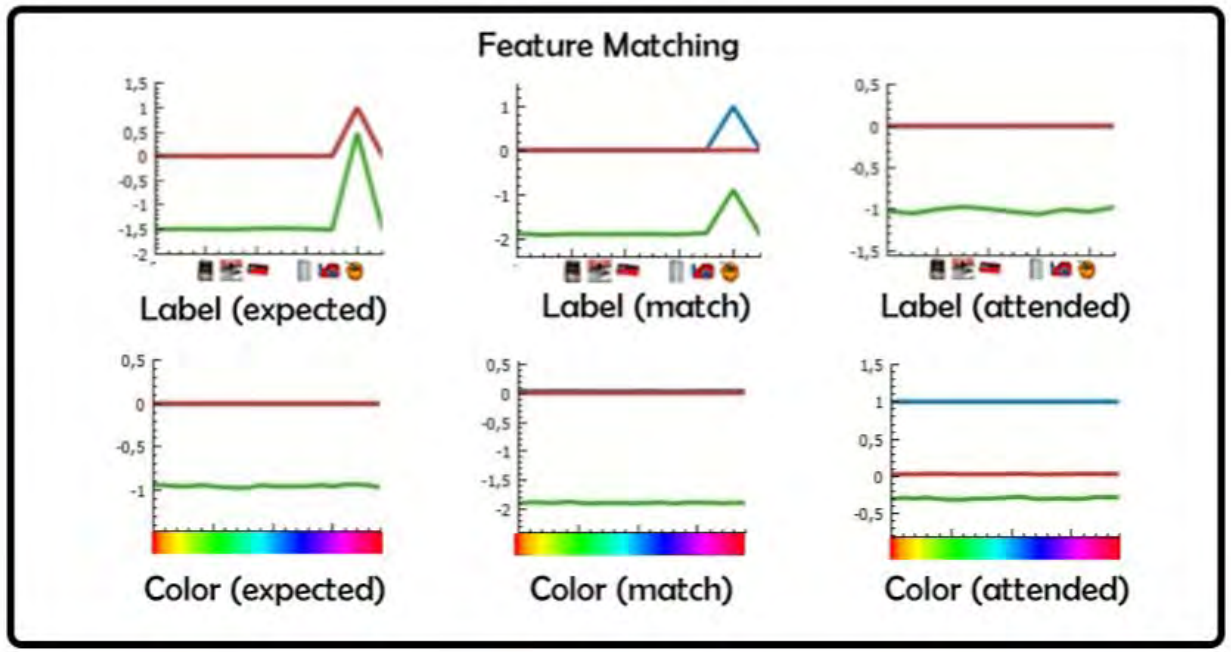
Attention (Sig. Activation)







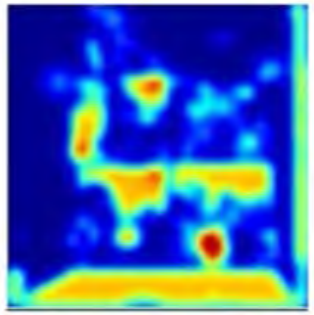




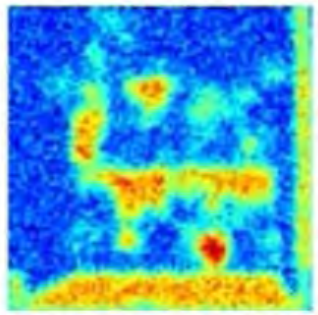
Camera Image

Foveal Image

Target Position (WM)

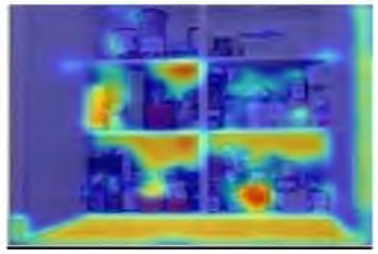


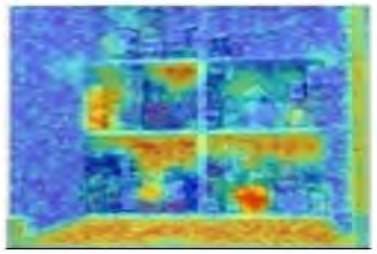
Attention (Input)

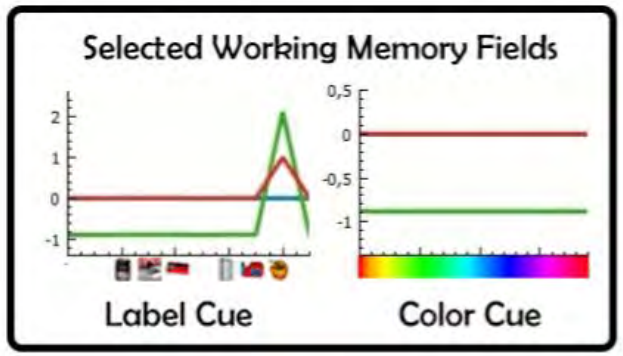


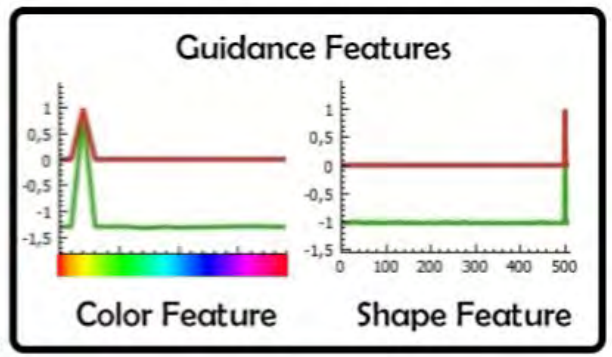
Attention (Activation)

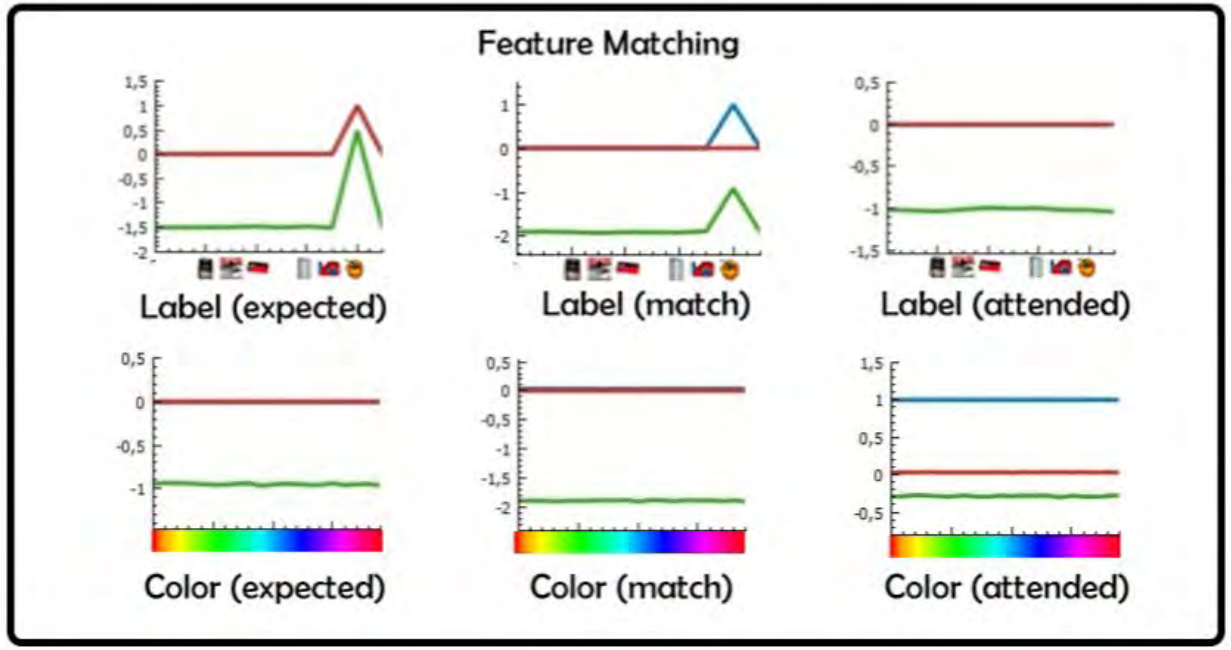
Attention (Sig. Activation)







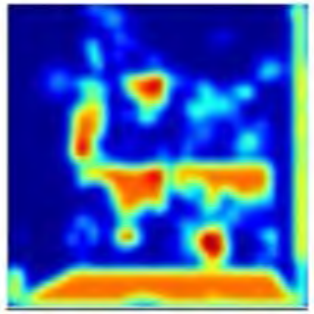




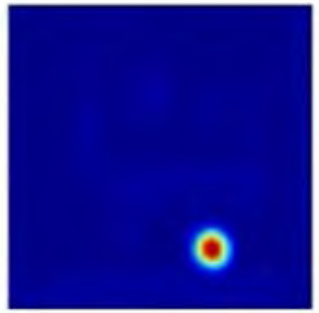
Camera Image

Foveal Image

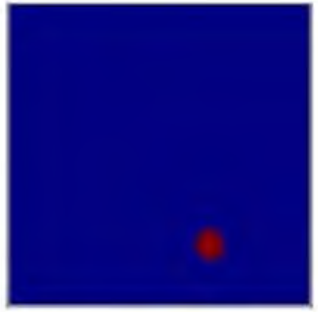
Target Position (WM)



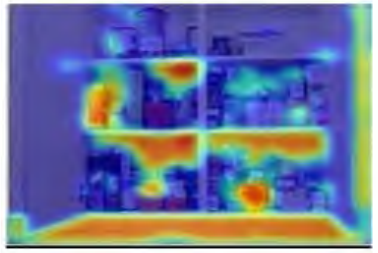
Attention (Input)

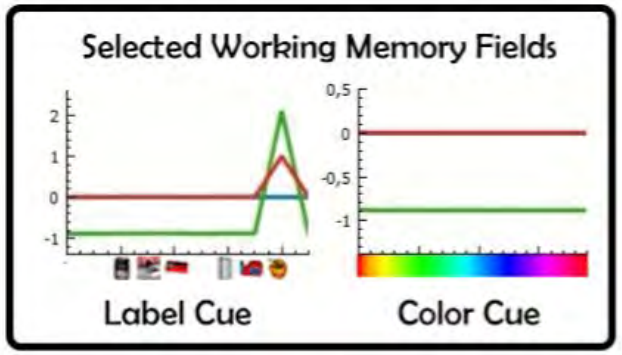


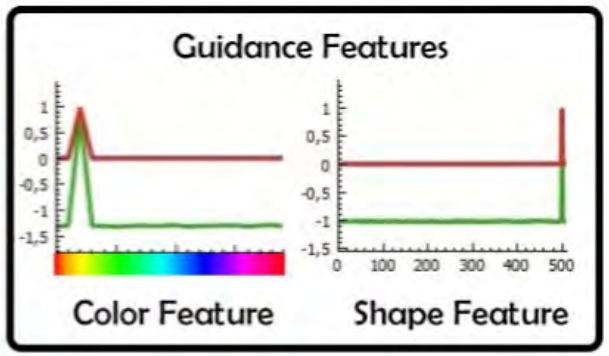
Attention (Activation)

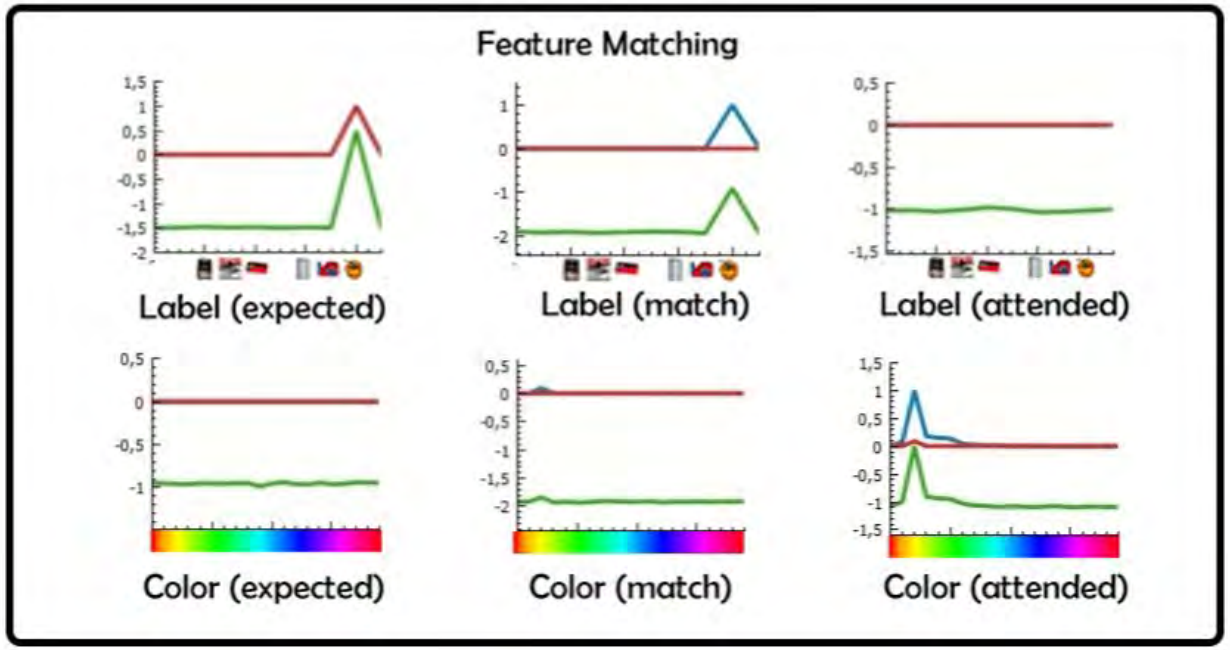


Attention (Sig. Activation)





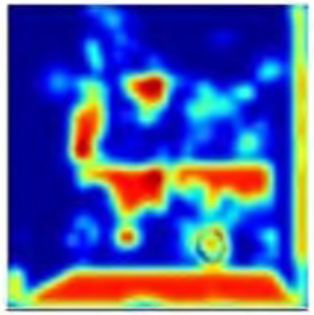




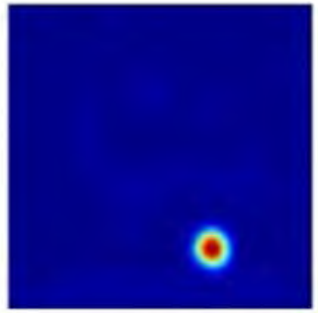
Camera Image

Foveal Image

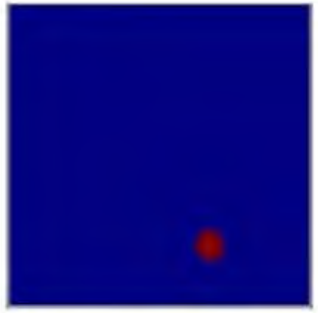
Target Position (WM)



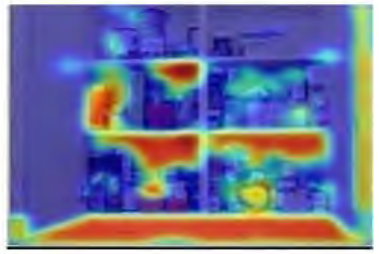
Attention (Input)

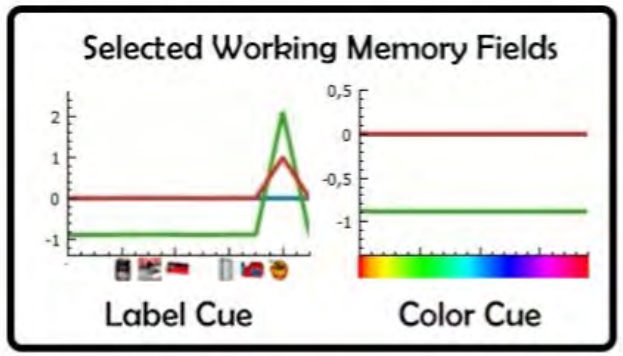


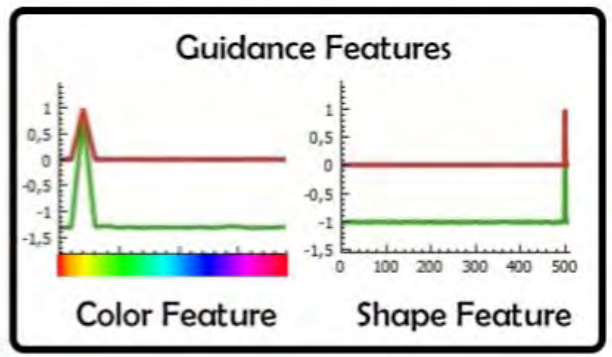
Attention (Activation)

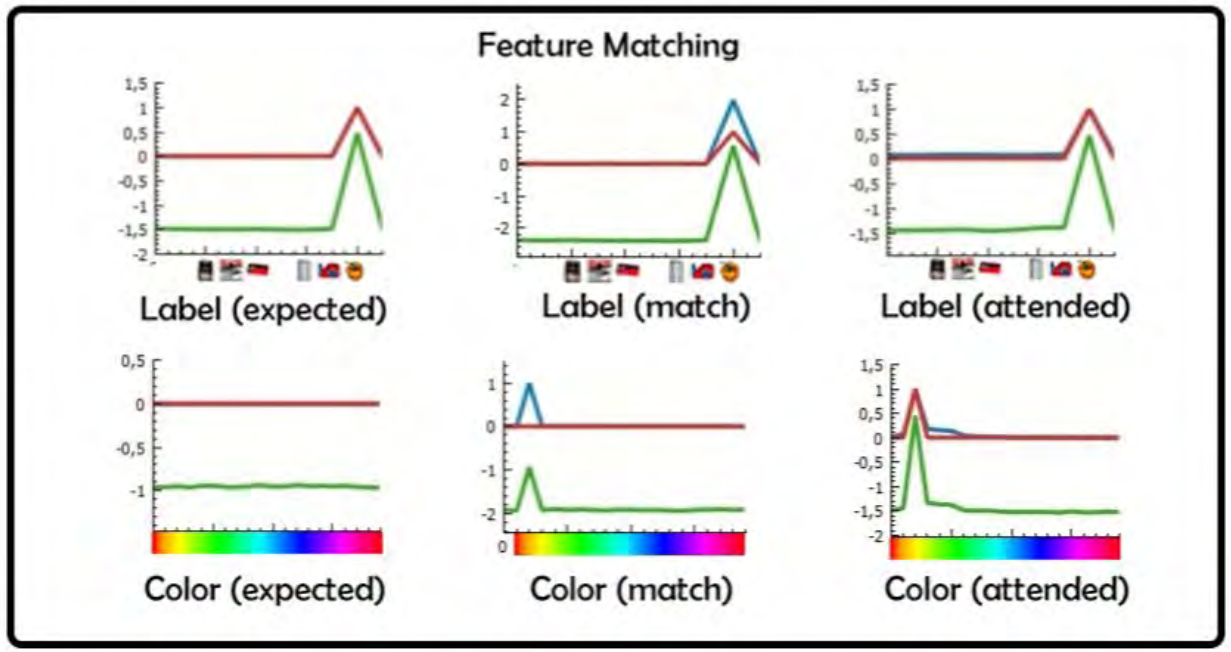


Attention (Sig. Activation)

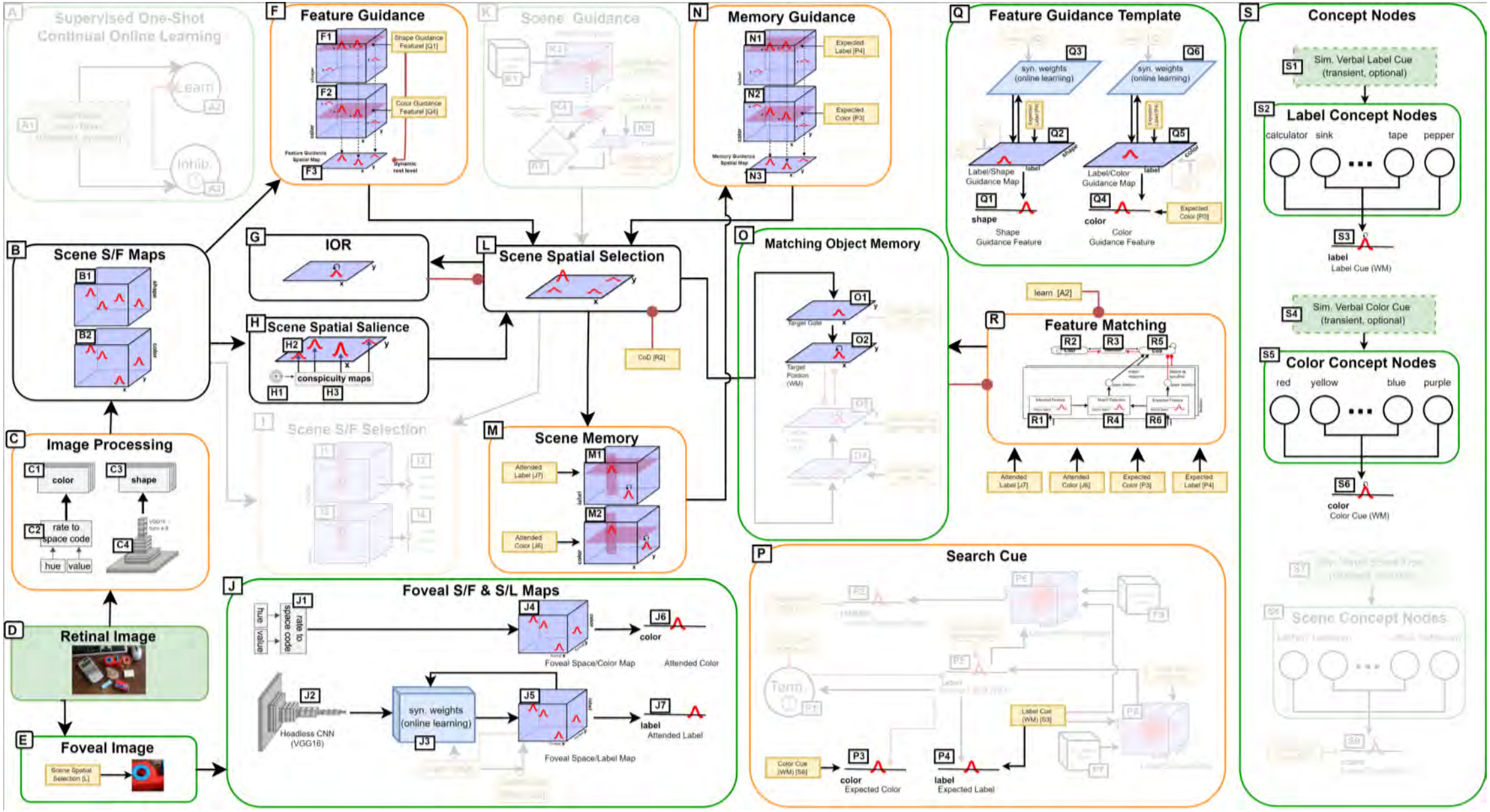








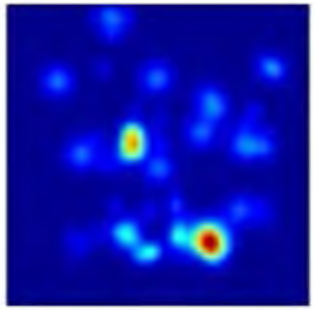
Combined categorical and basic feature visual search



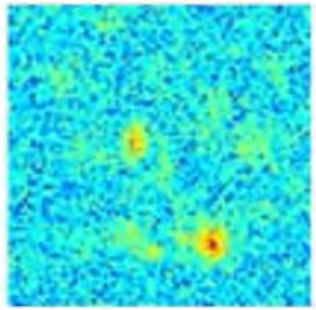
Camera Image

Foveal Image

Target Position (WM)

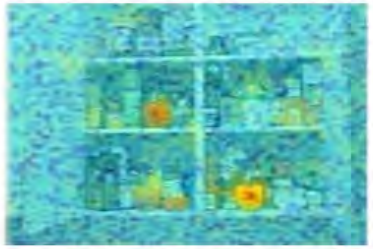


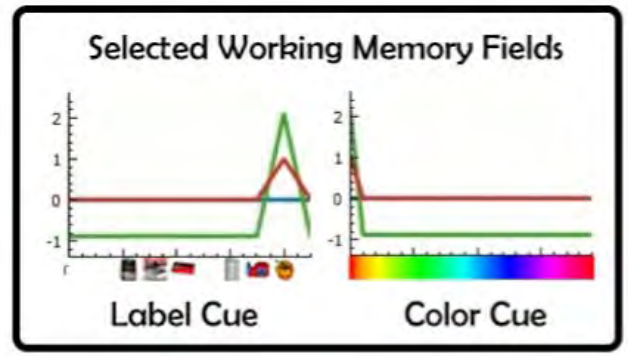
Attention (Input)

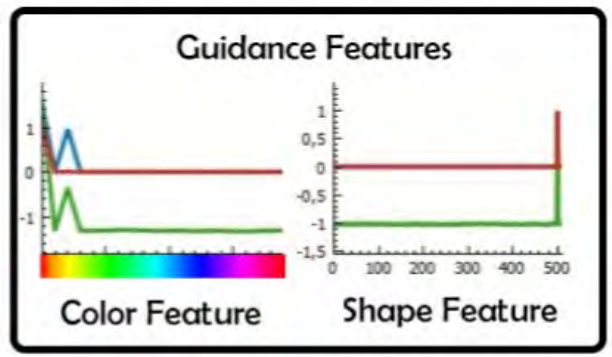


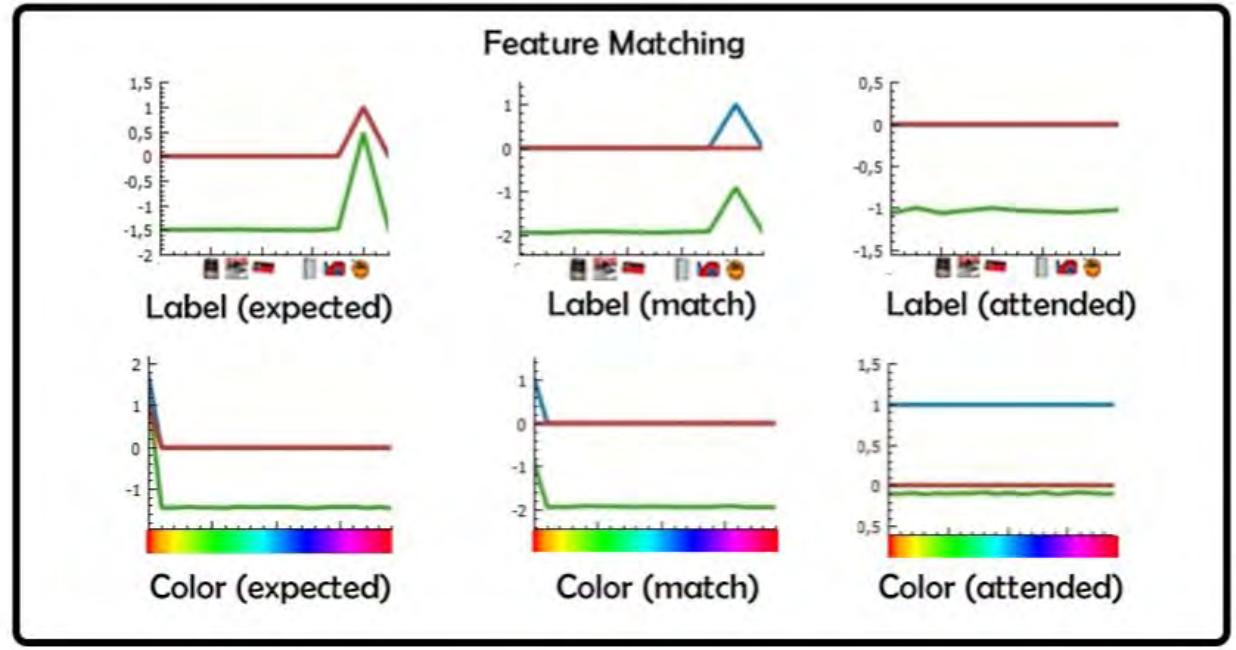
Attention (Activation)

Attention (Sig. Activation)





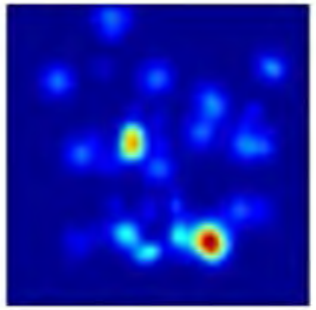




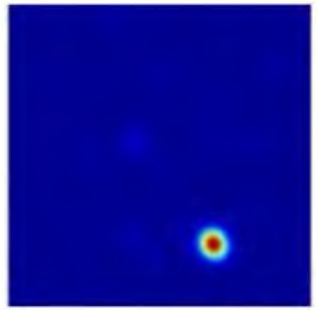
Camera Image

Foveal Image

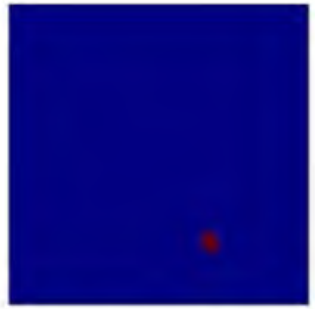
Target Position (WM)



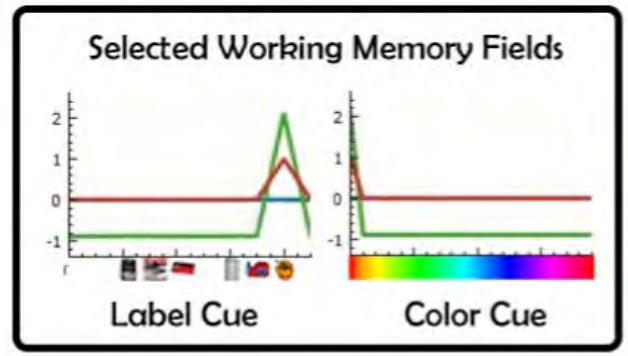
Attention (Input)

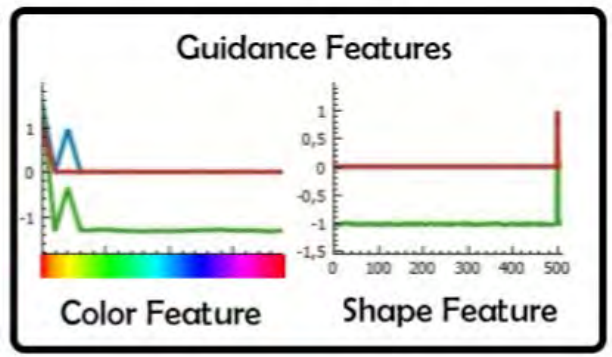


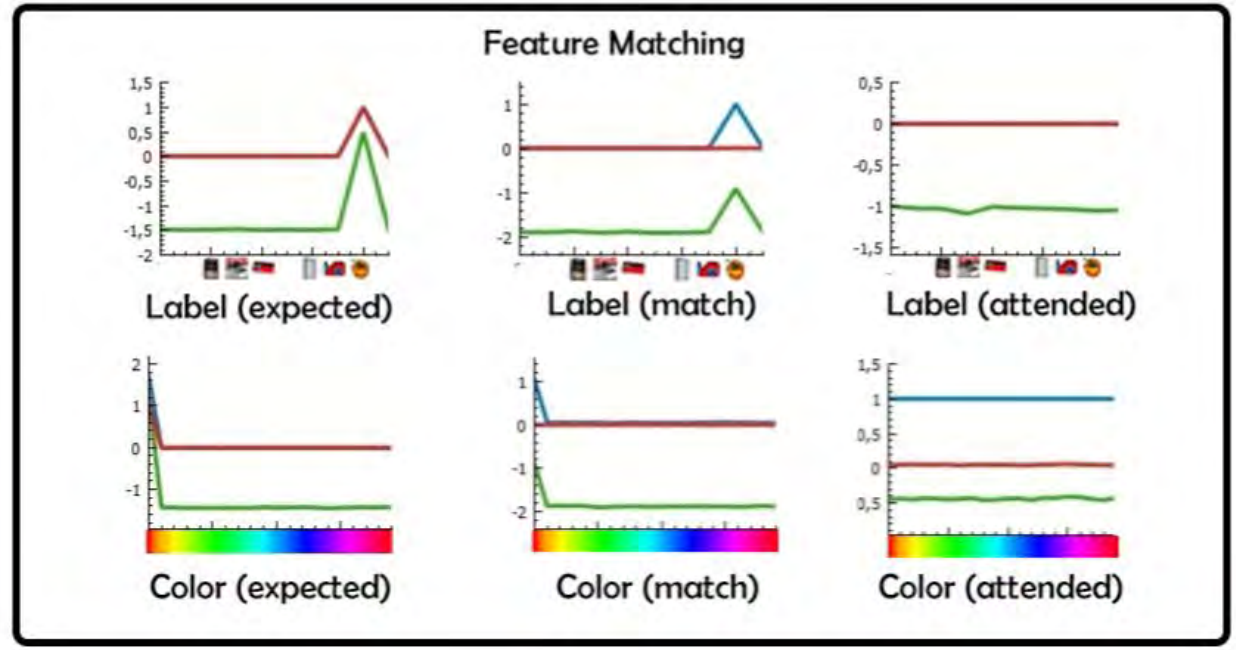
Attention (Activation)



Attention (Sig. Activation)



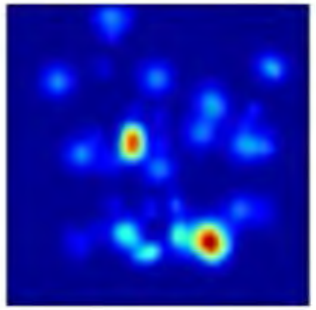




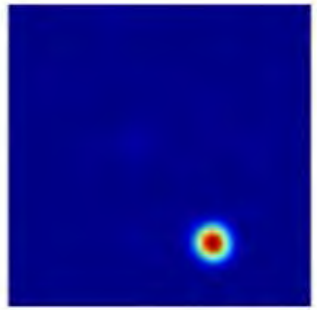
Camera Image

Foveal Image

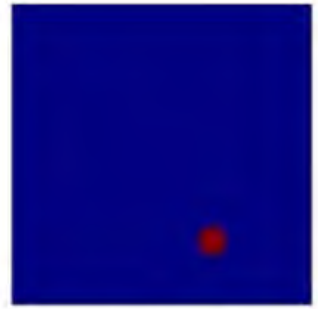
Target Position (WM)



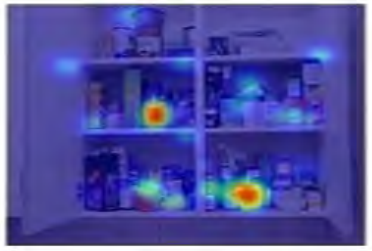
Attention (Input)

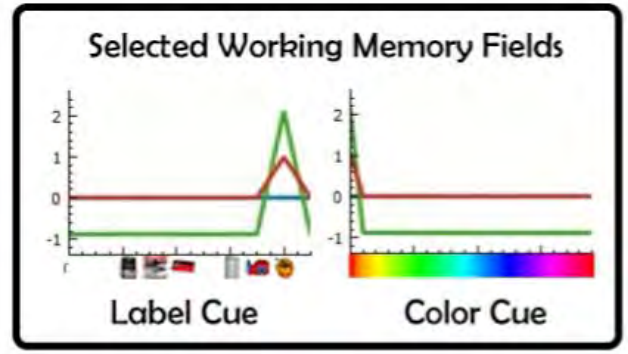


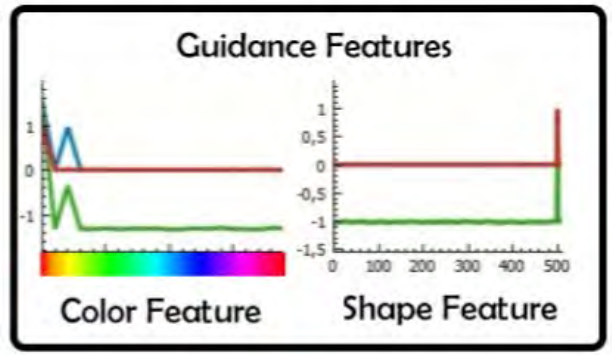
Attention (Activation)

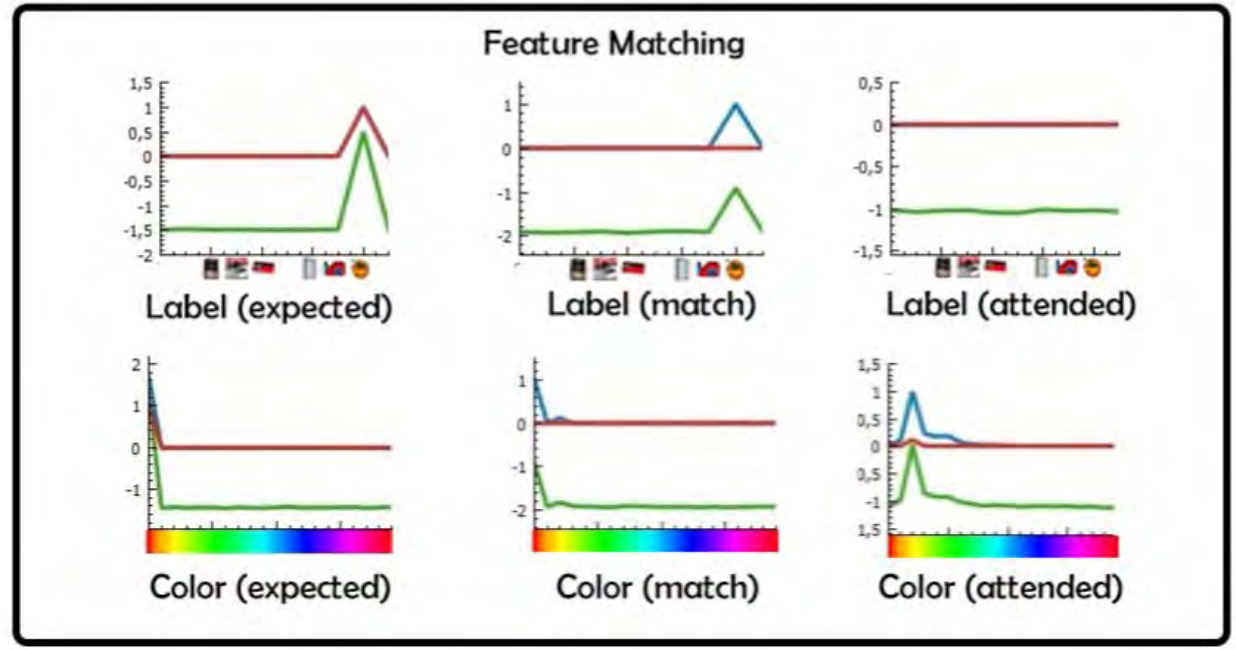


Attention (Sig. Activation)





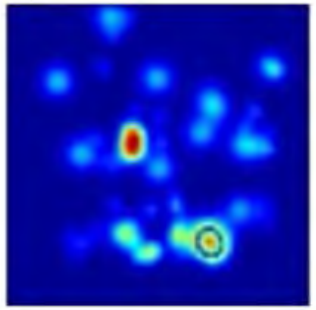




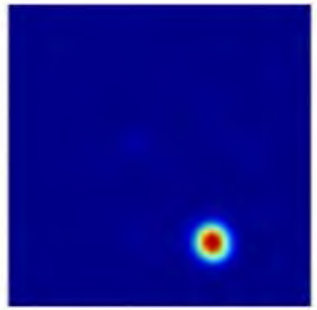
Camera Image

Foveal Image

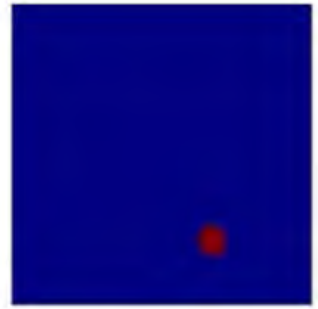
Target Position (WM)



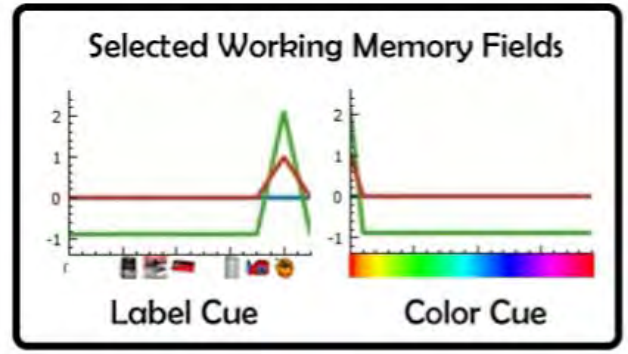
Attention (Input)

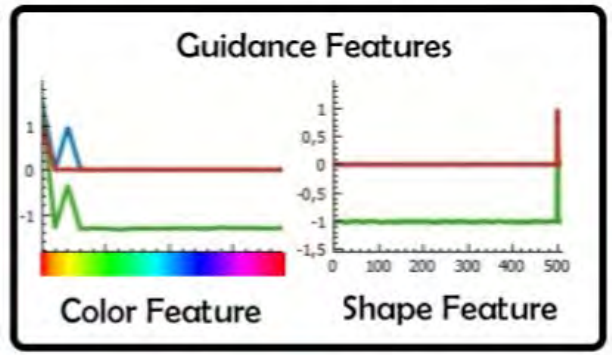


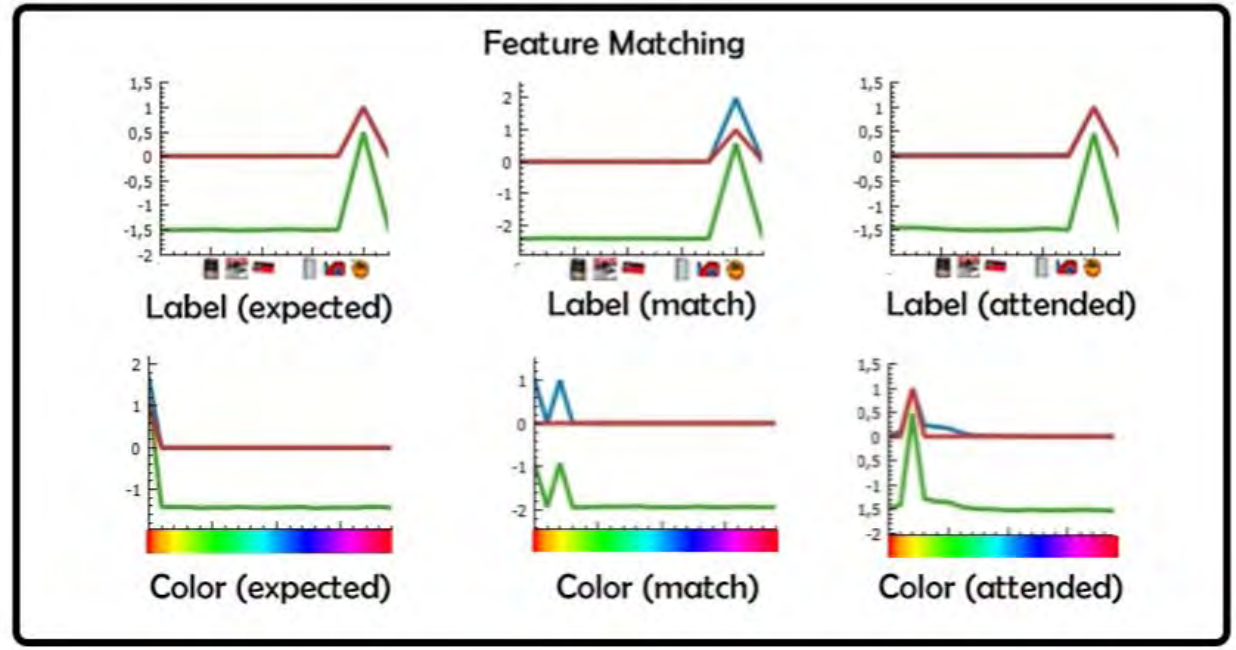
Attention (Activation)



Attention (Sig. Activation)



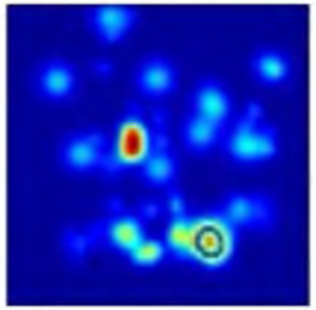




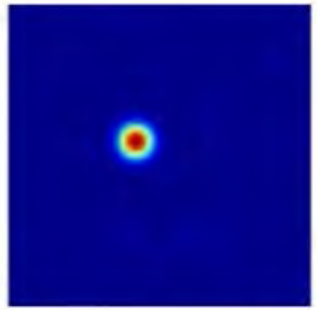
Camera Image

Foveal Image

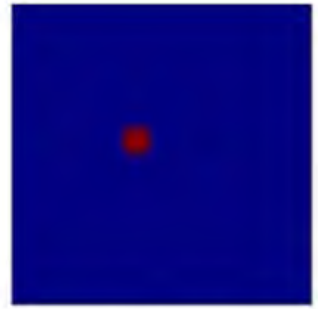
Target Position (WM)



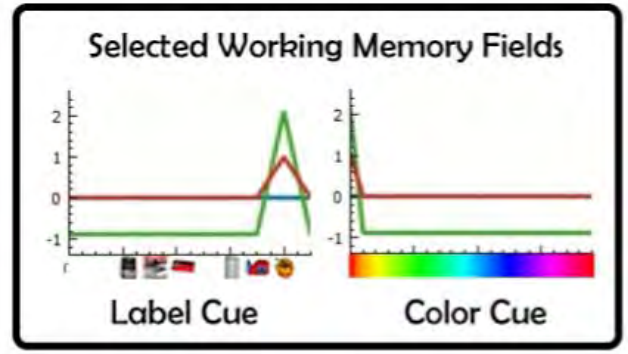
Attention (Input)

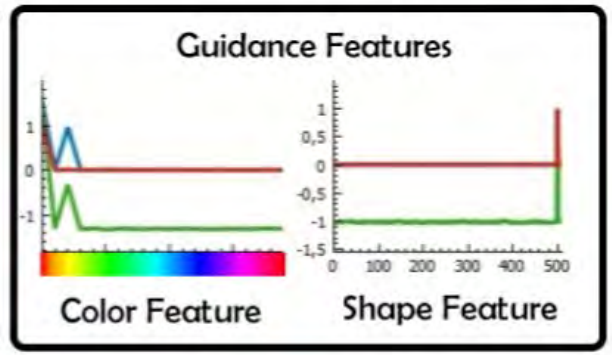


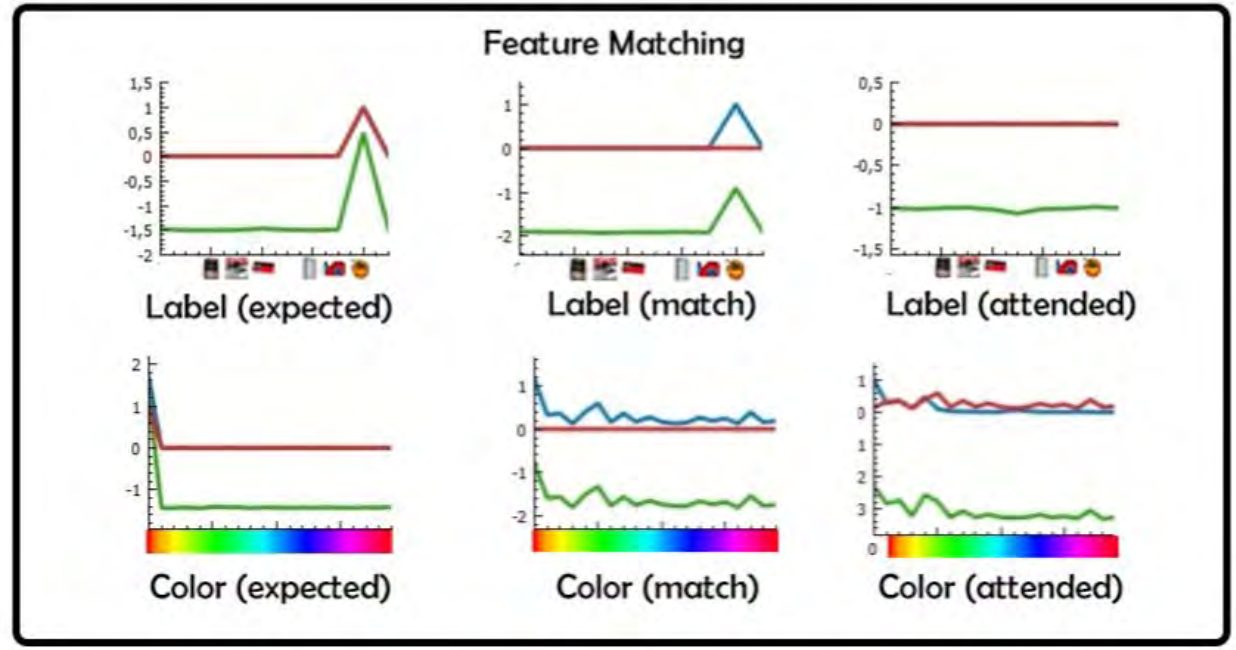
Attention (Activation)



Attention (Sig. Activation)



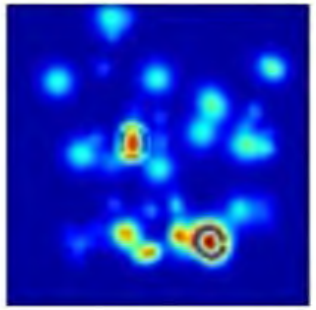




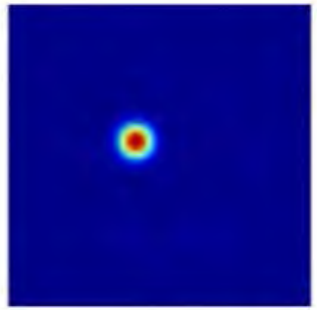
Camera Image

Foveal Image

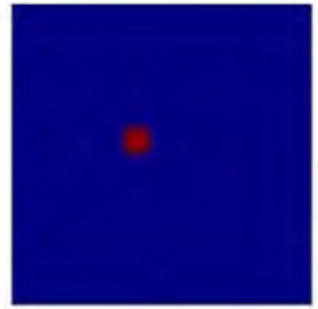
Target Position (WM)



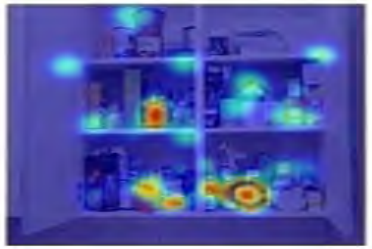
Attention (Input)

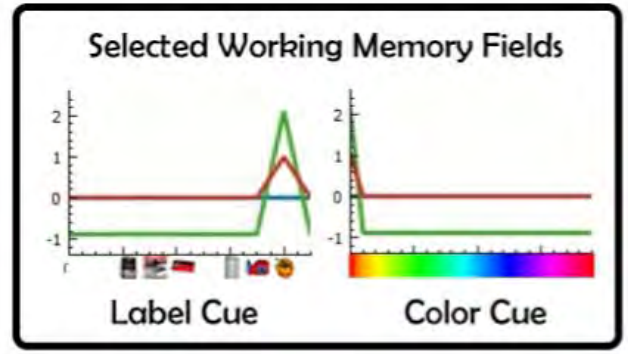


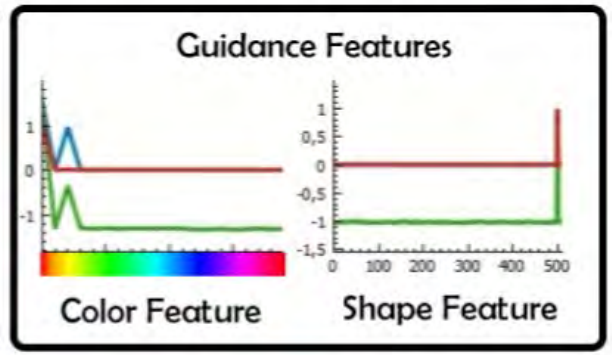
Attention (Activation)

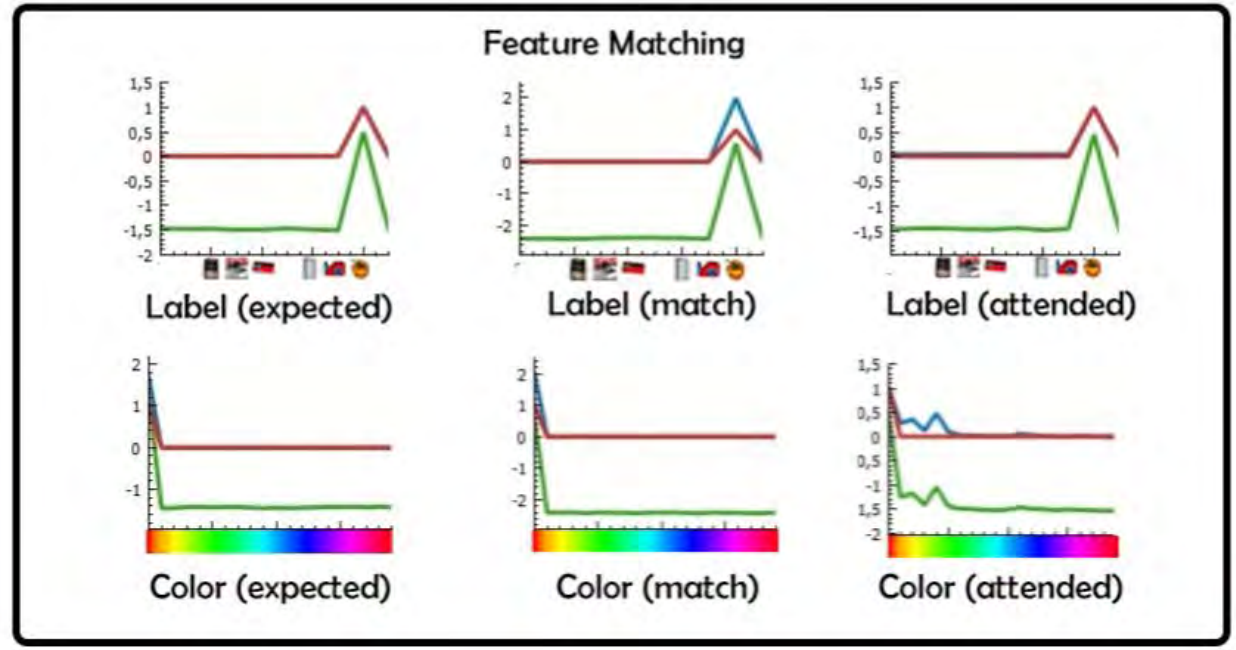


Attention (Sig. Activation)

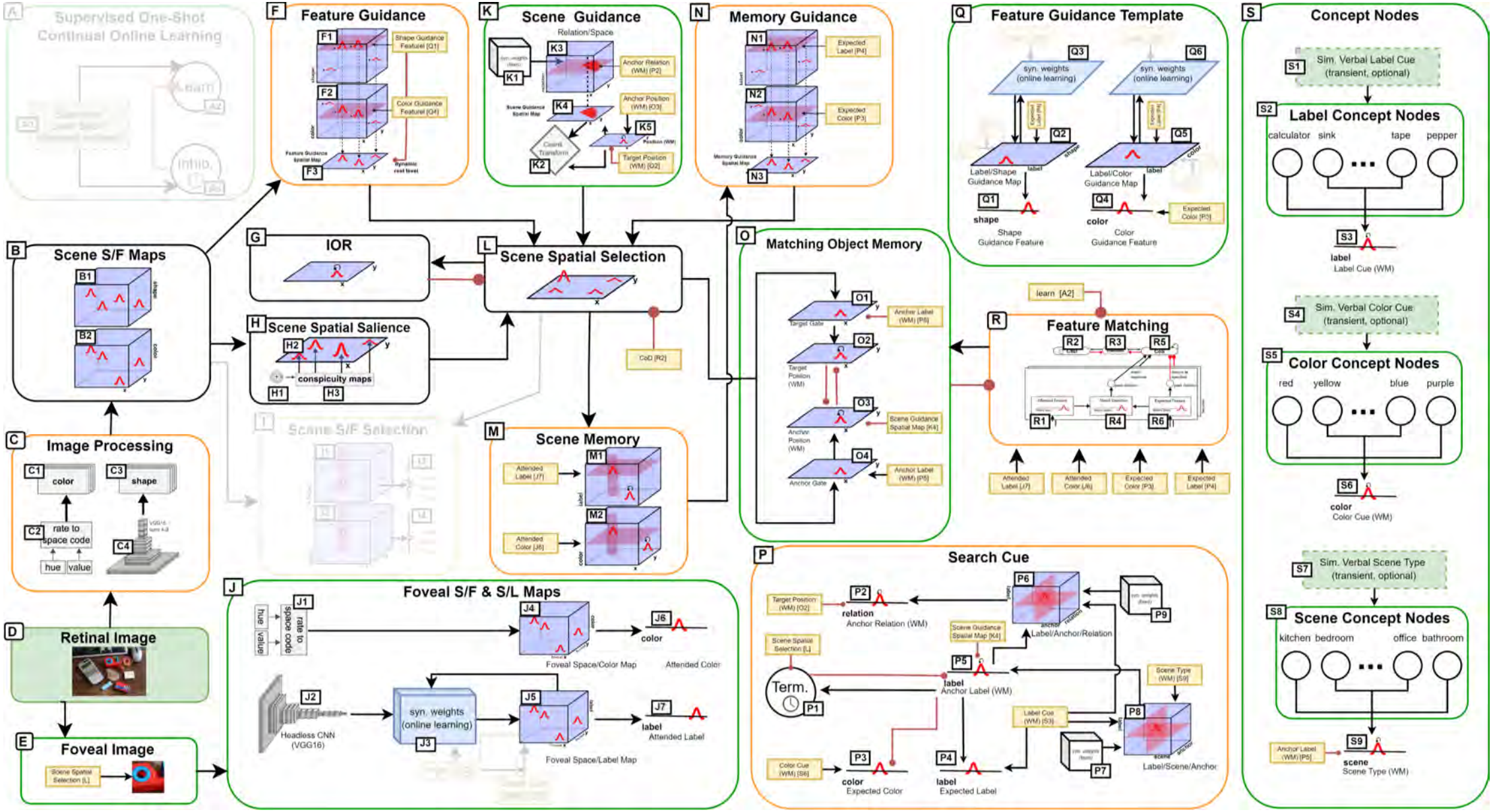








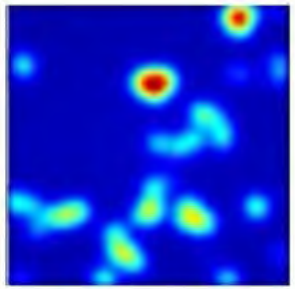
Scene grammar



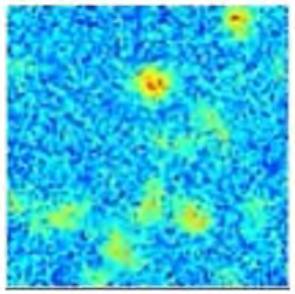
Camera Image

Foveal Image

Anchor Position (WM)

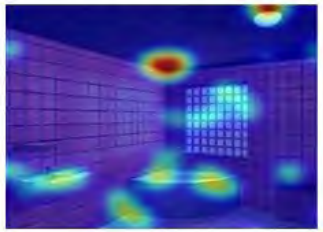


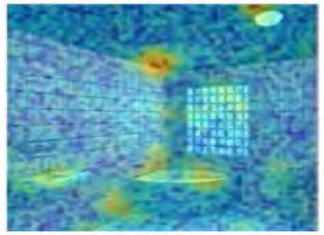
Attention (Input)

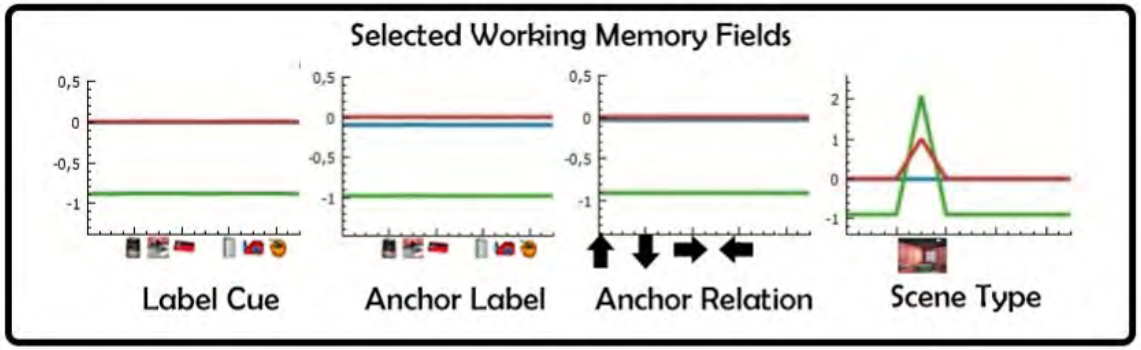


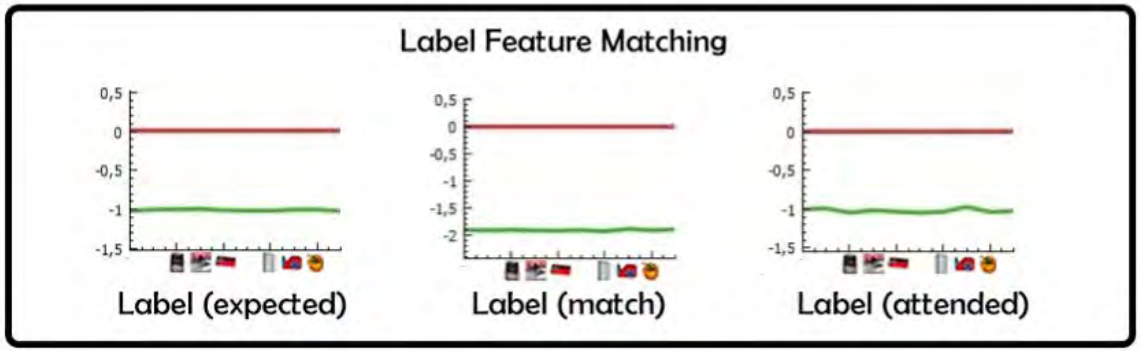
Attention (Activation)

Attention (Sig. Activation)





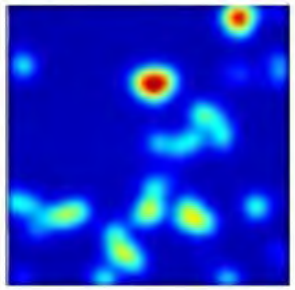




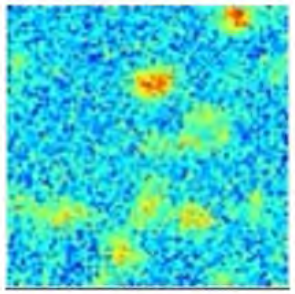
Camera Image

Foveal Image

Anchor Position (WM)

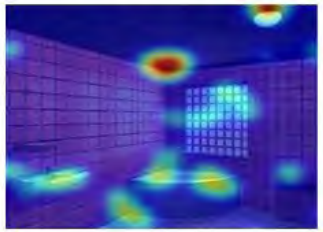


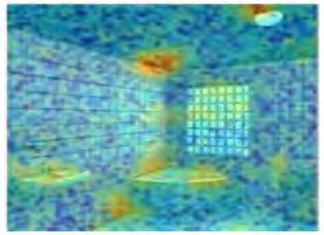
Attention (Input)

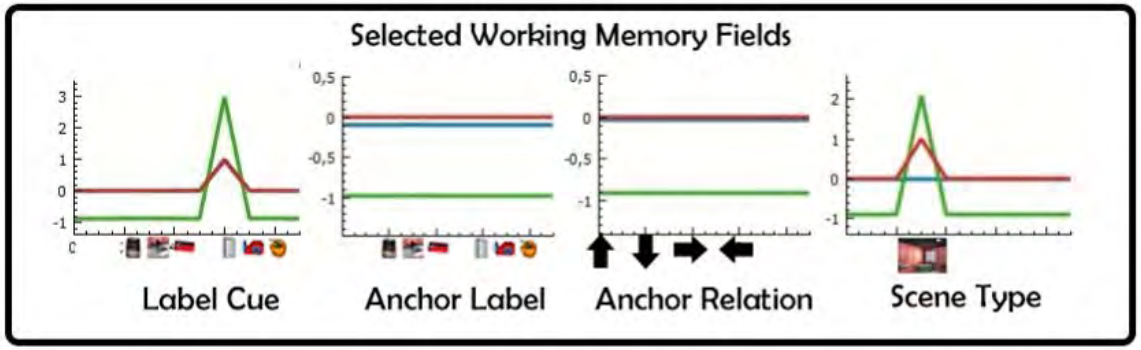


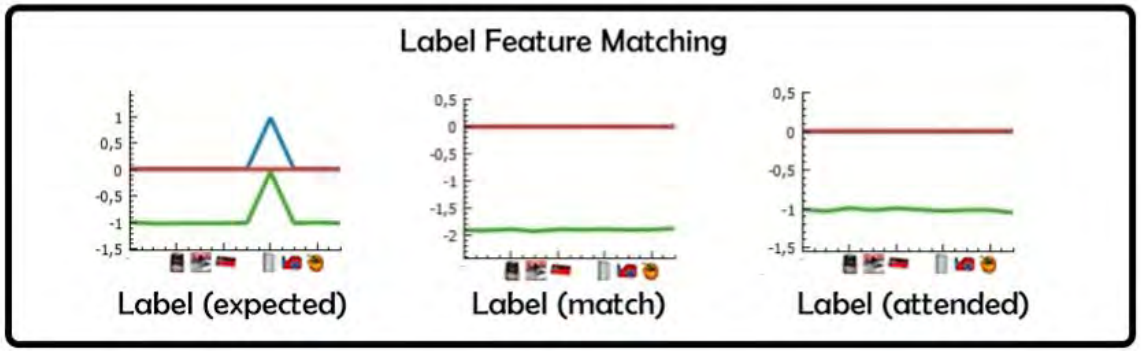
Attention (Activation)

Attention (Sig. Activation)





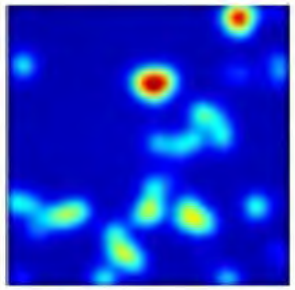




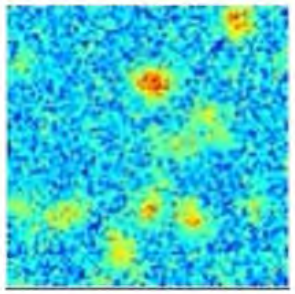
Camera Image

Foveal Image

Anchor Position (WM)

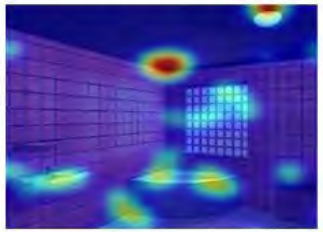


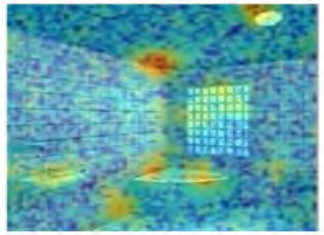
Attention (Input)

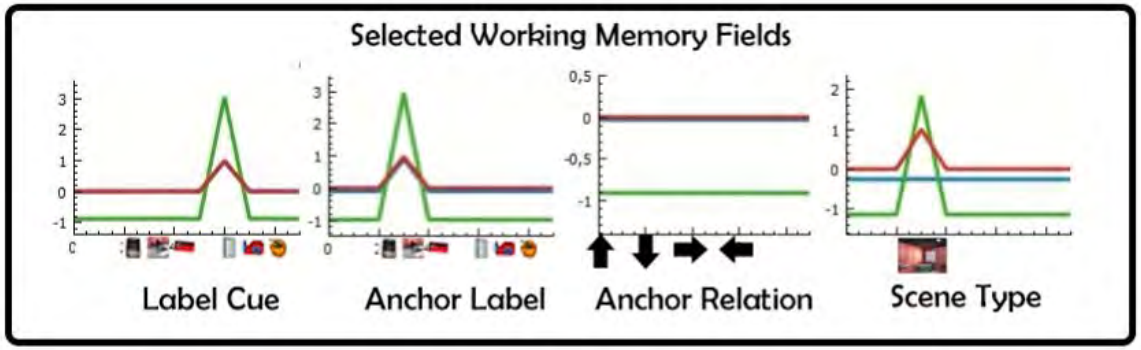


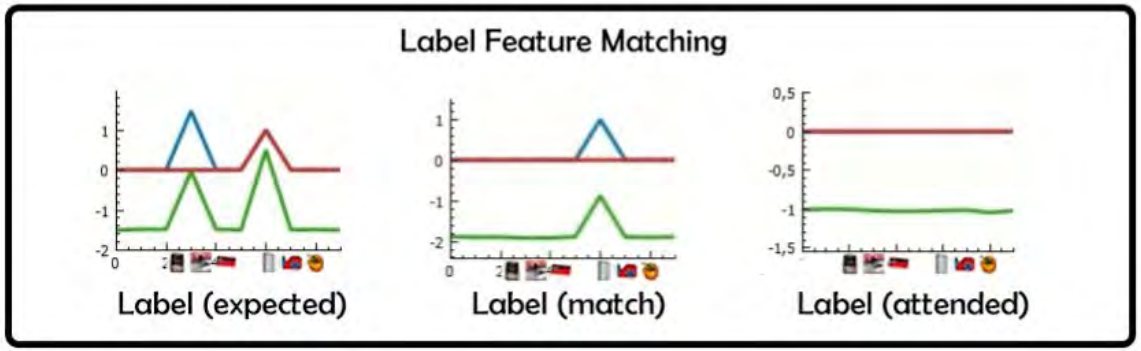
Attention (Activation)

Attention (Sig. Activation)





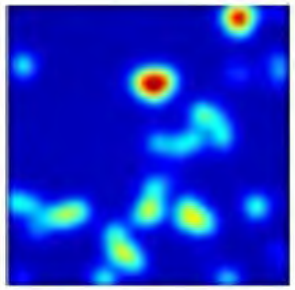




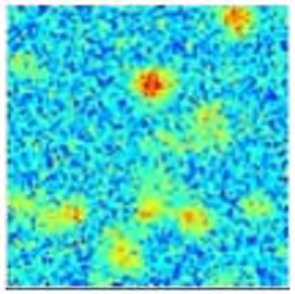
Camera Image

Foveal Image

Anchor Position (WM)

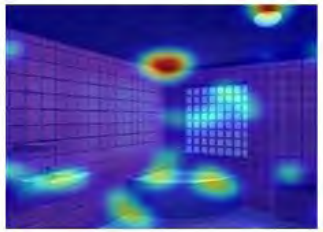


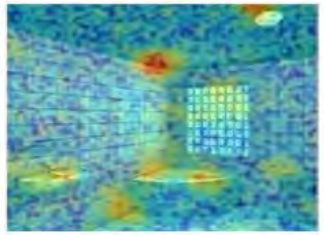
Attention (Input)

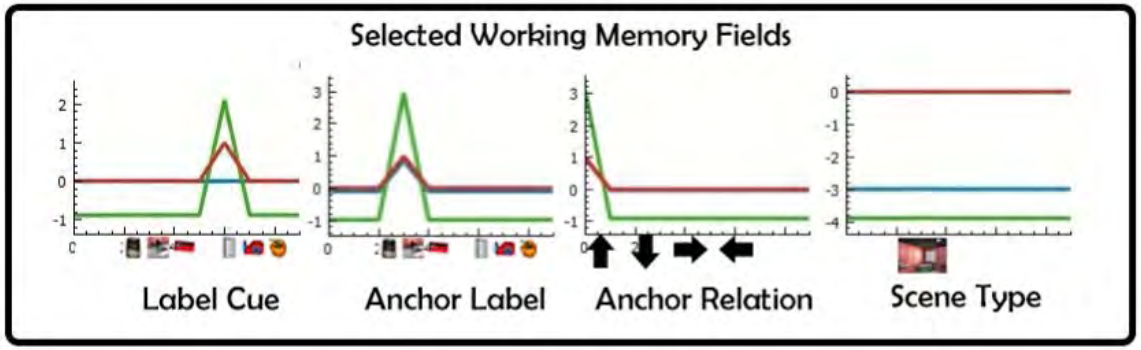


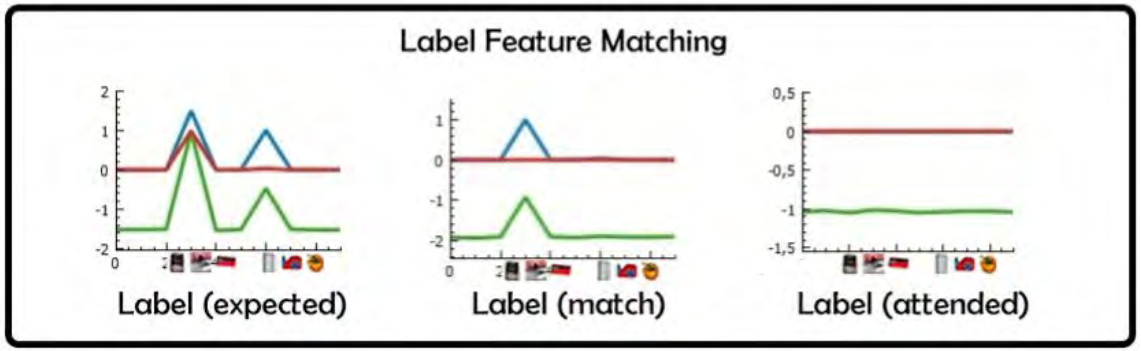
Attention (Activation)

Attention (Sig. Activation)





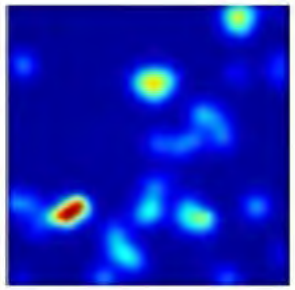




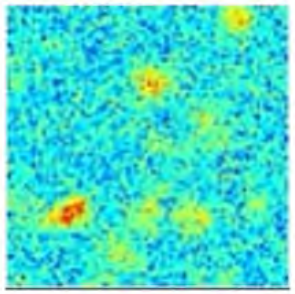
Camera Image

Foveal Image

Anchor Position (WM)

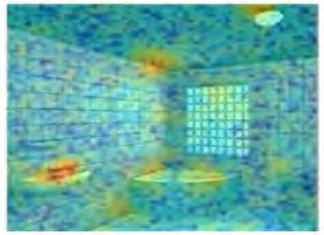


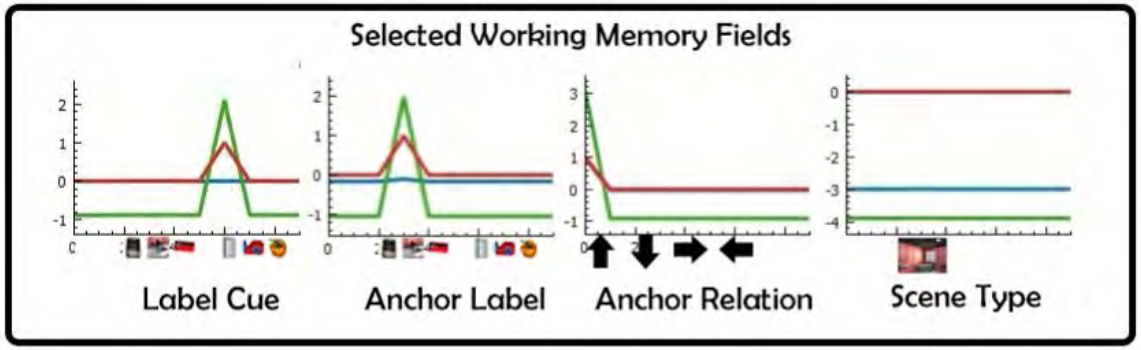
Attention (Input)

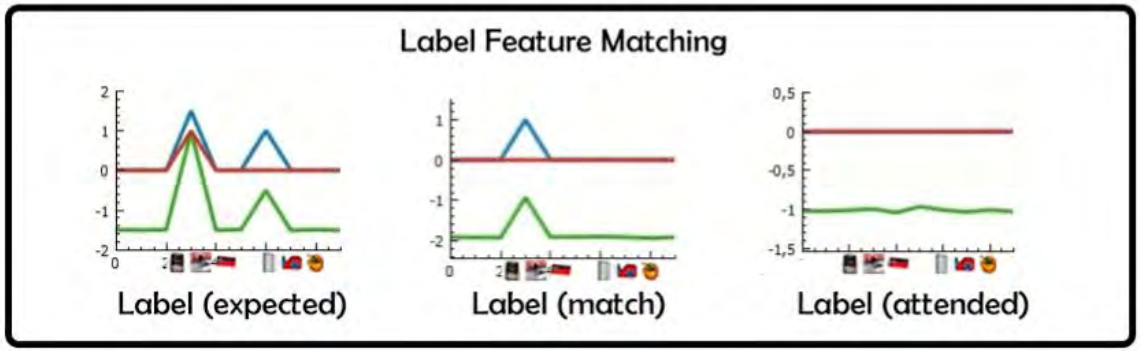


Attention (Activation)

Attention (Sig. Activation)



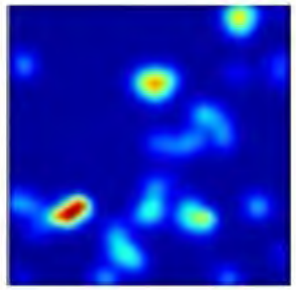




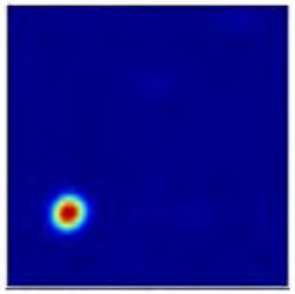
Camera Image

Foveal Image

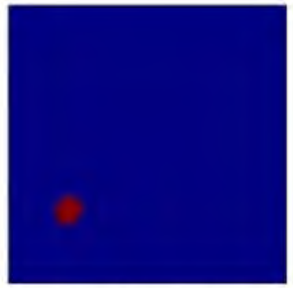
Anchor Position (WM)



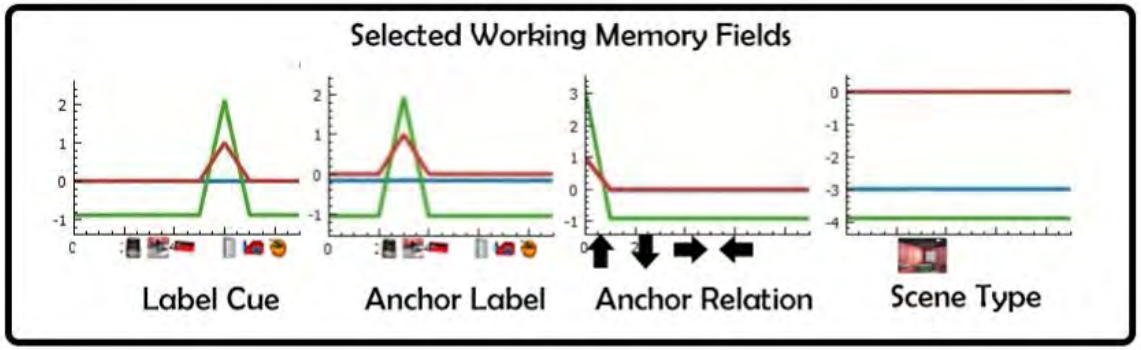
Attention (Input)

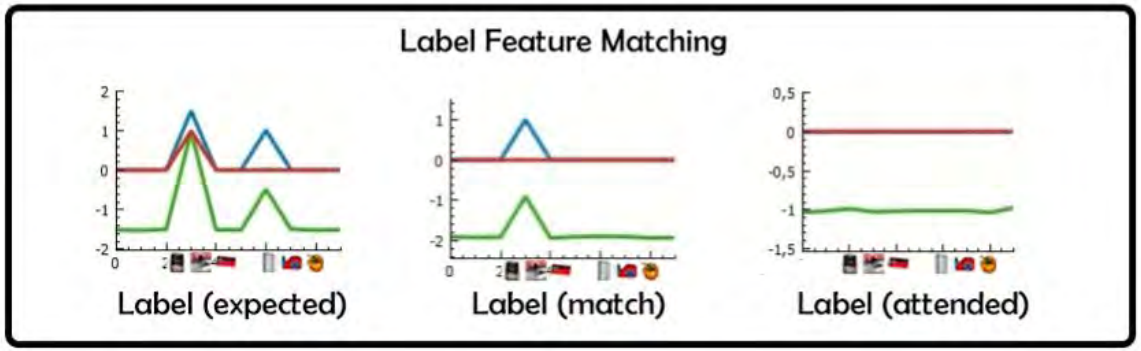


Attention (Activation)



Attention (Sig. Activation)

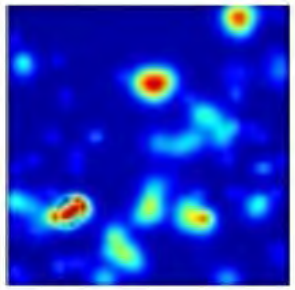




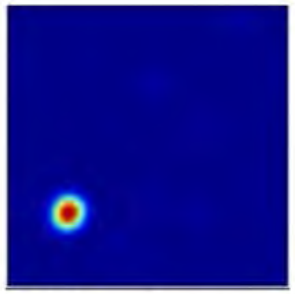
Camera Image

Foveal Image

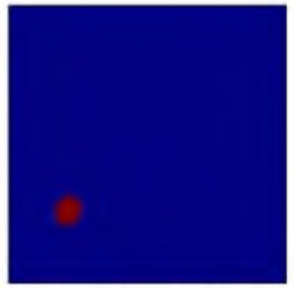
Anchor Position (WM)



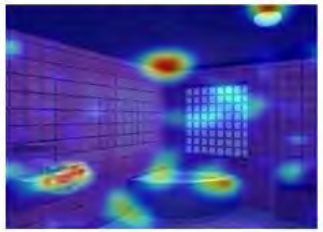
Attention (Input)

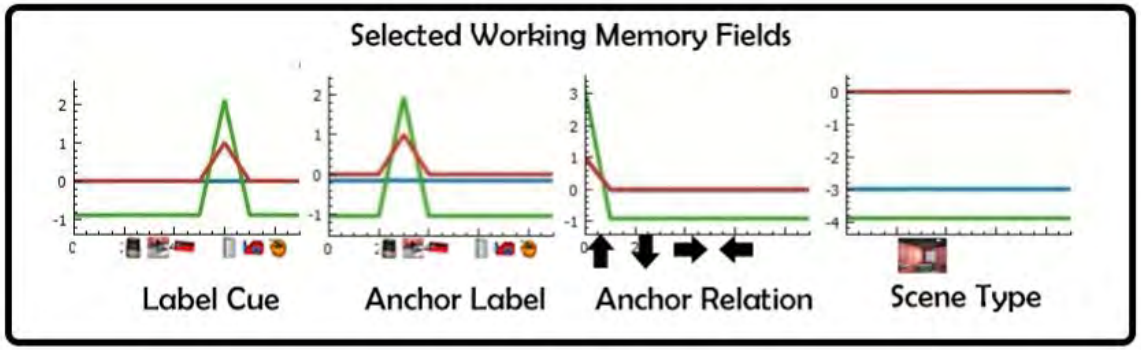


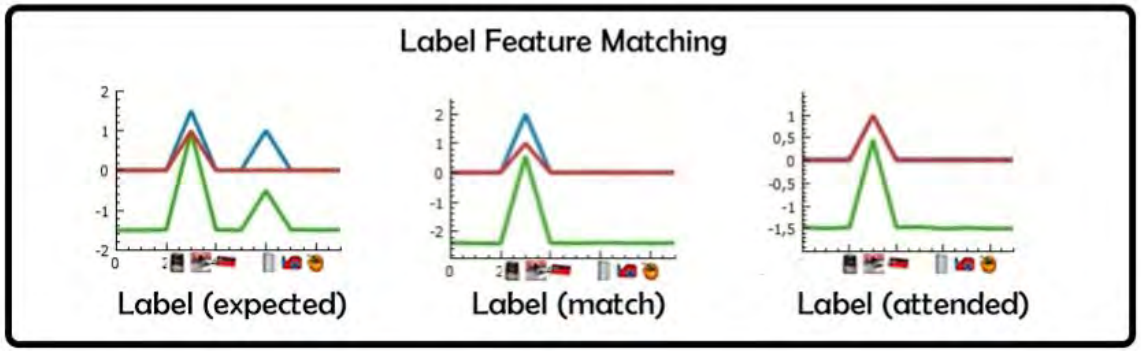
Attention (Activation)



Attention (Sig. Activation)

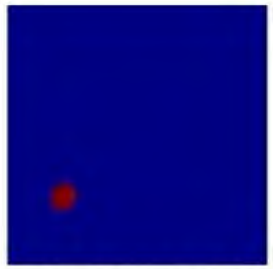




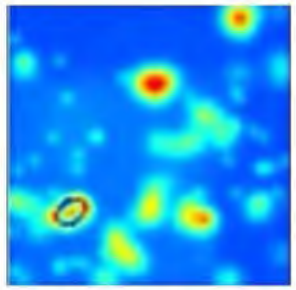


Camera Image

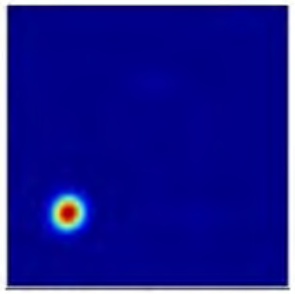
Foveal Image



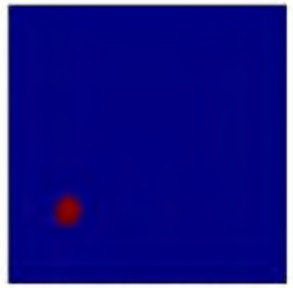
Anchor Position (WM)



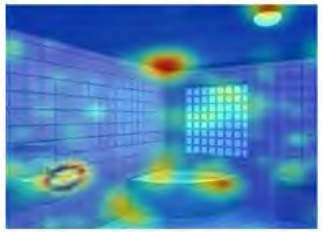
Attention (Input)

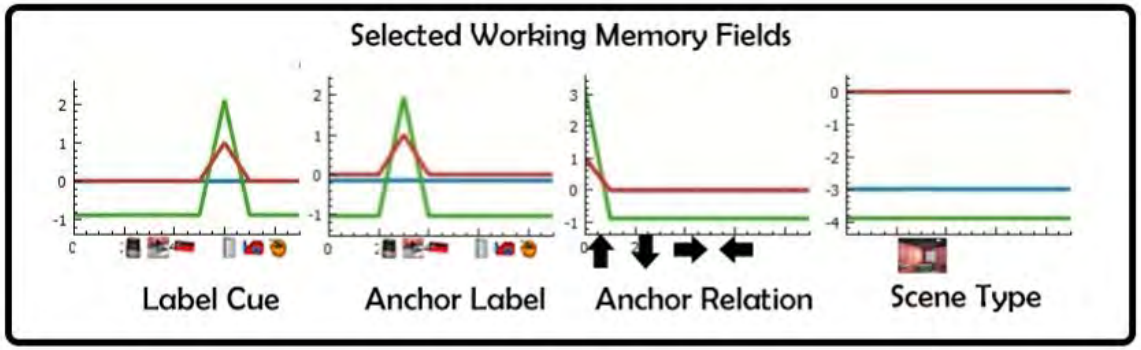


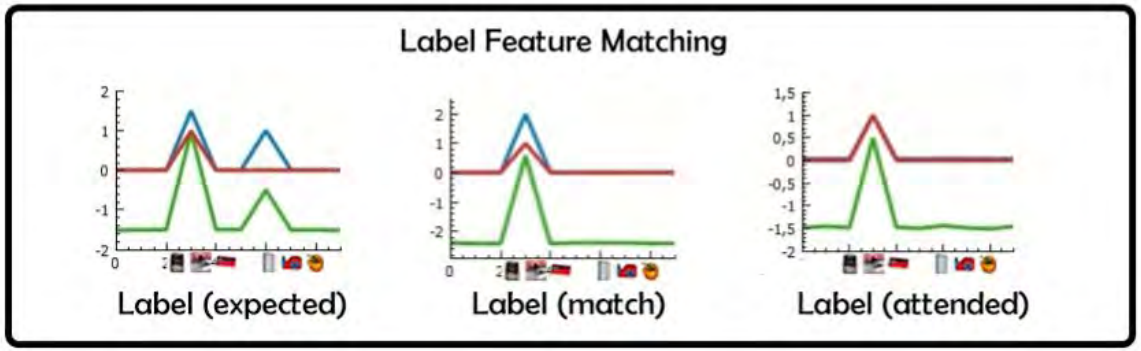
Attention (Activation)



Attention (Sig. Activation)

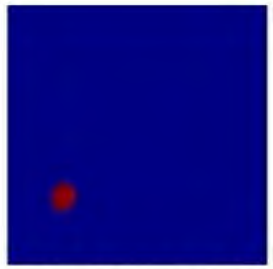




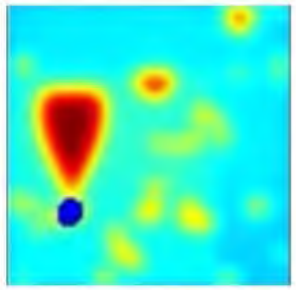


Camera Image

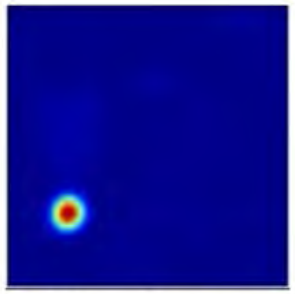
Foveal Image



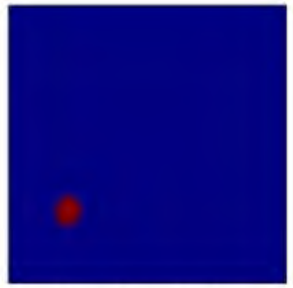
Anchor Position (WM)



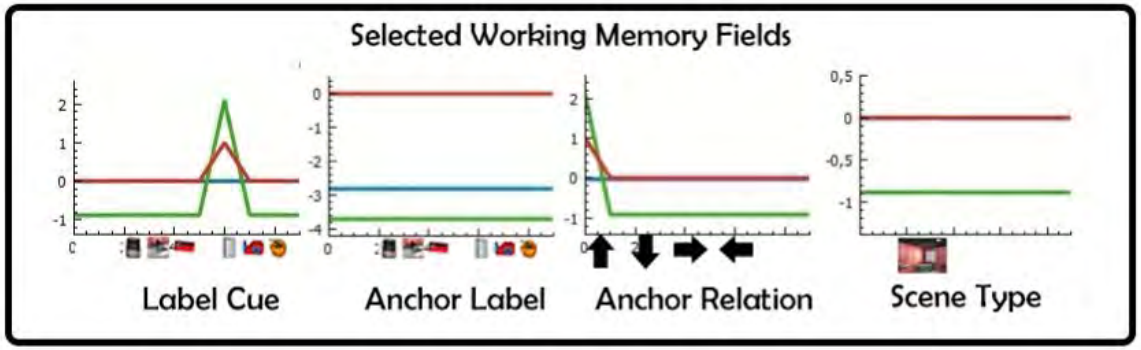
Attention (Input)

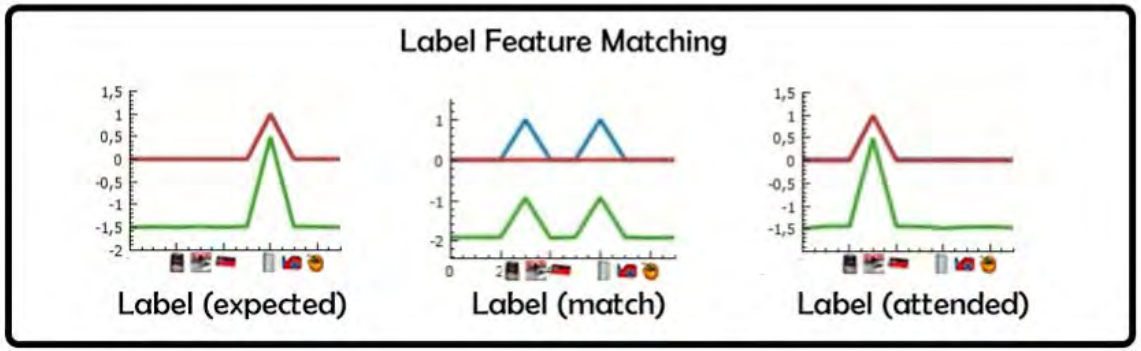


Attention (Activation)



Attention (Sig. Activation)





- We have shown a neural process account of visual search and scene memory that autonomously builds a scene representation and performs guided categorical visual search on natural scenes

- We found solutions for three important problems:

  - How the association between a categorical concept and preattentive shape can be learned from the intermediate layer of a CNN

  - How the distributed representation of the CNN feature maps can be mapped to the localist representation of a dynamic neural field

  - How scene grammar emerges from the underlying neural dynamics