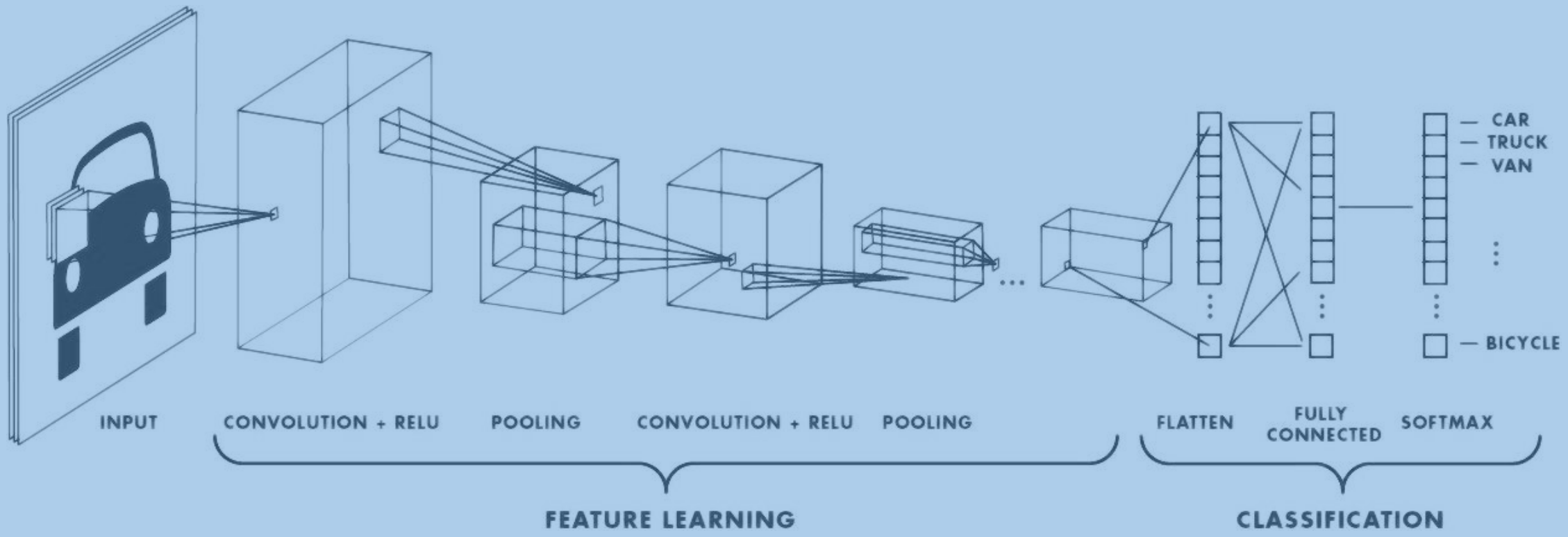
Bridging DFT and DNNs:

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

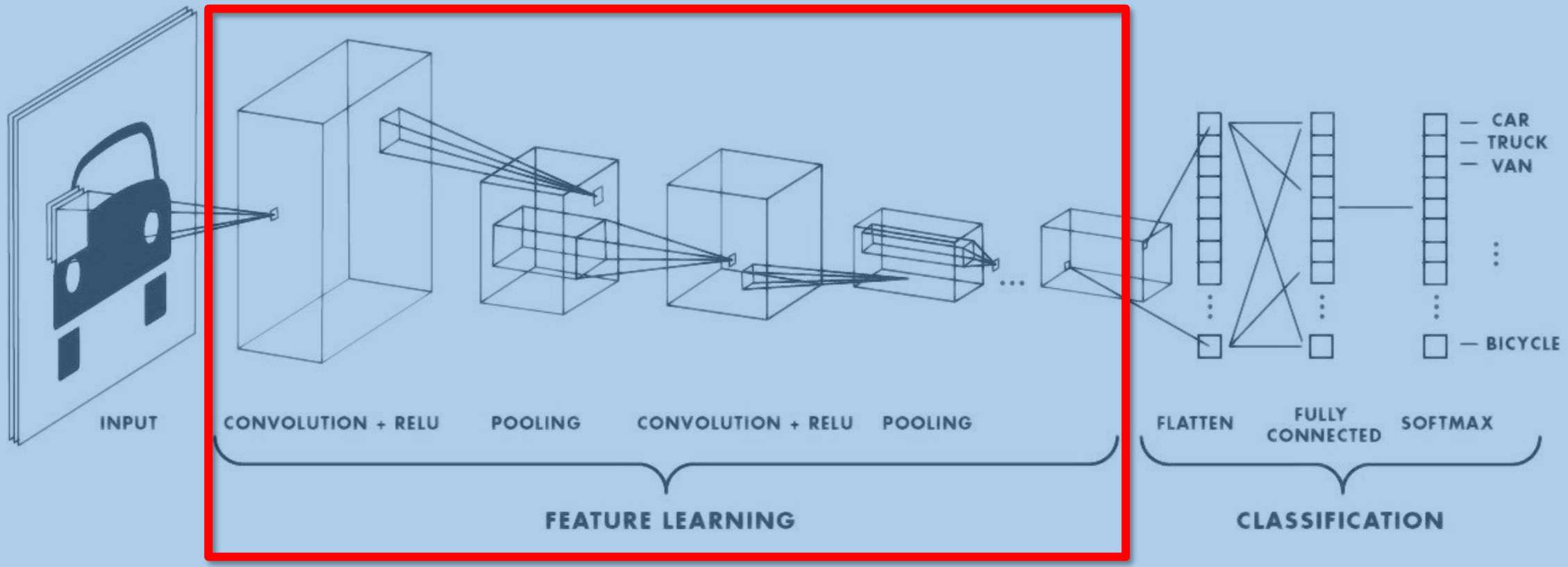
Raul Grieben

24.08.2023

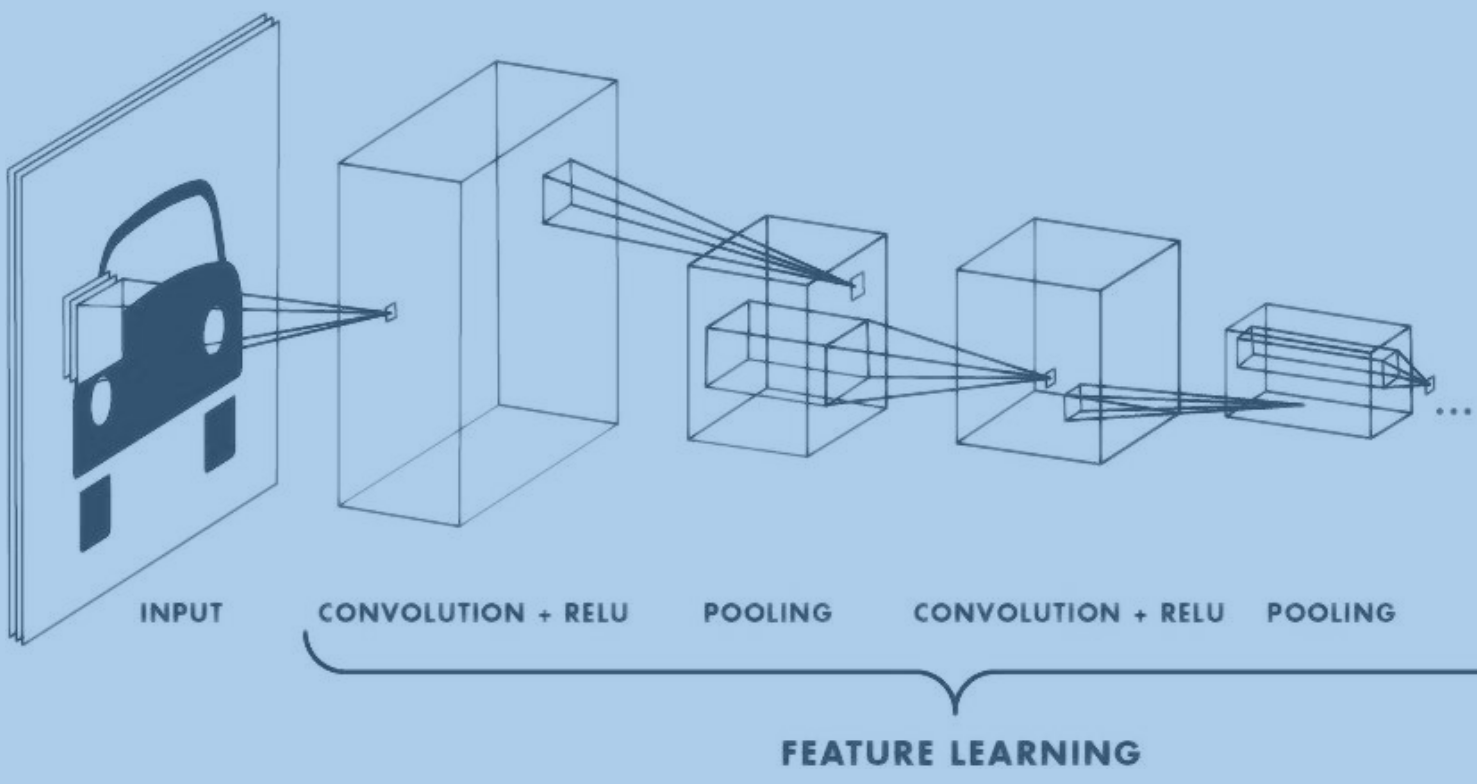
Deep CNN


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

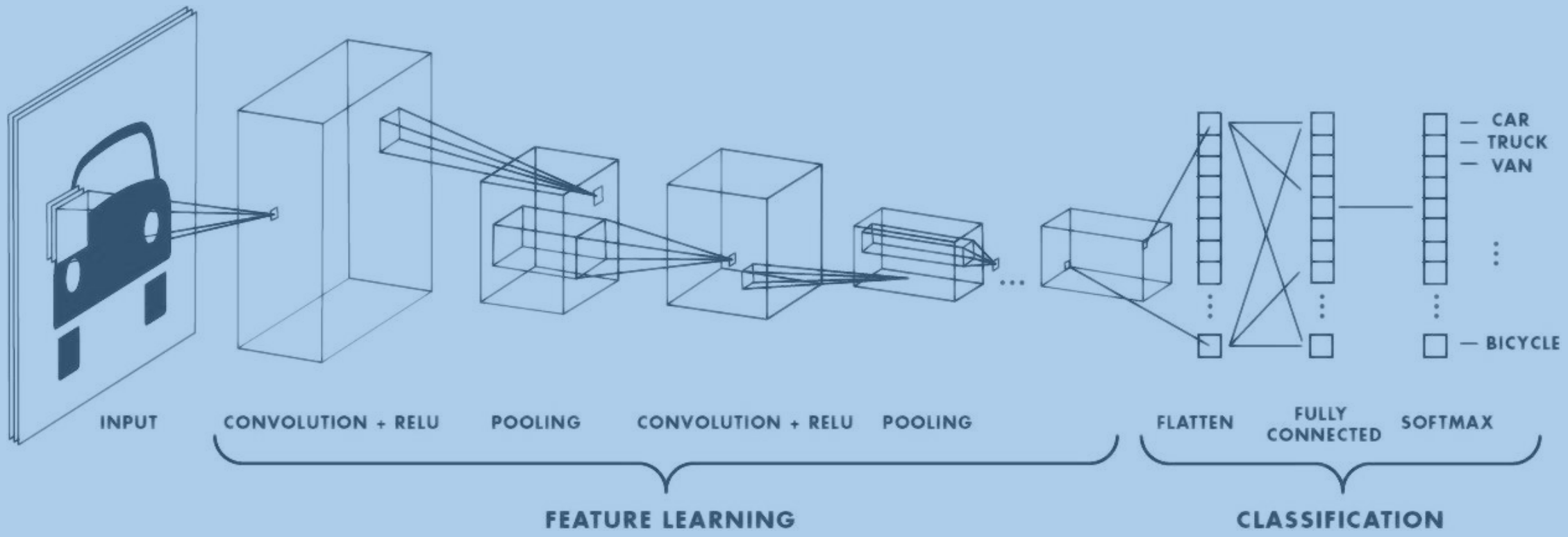
Deep CNN


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

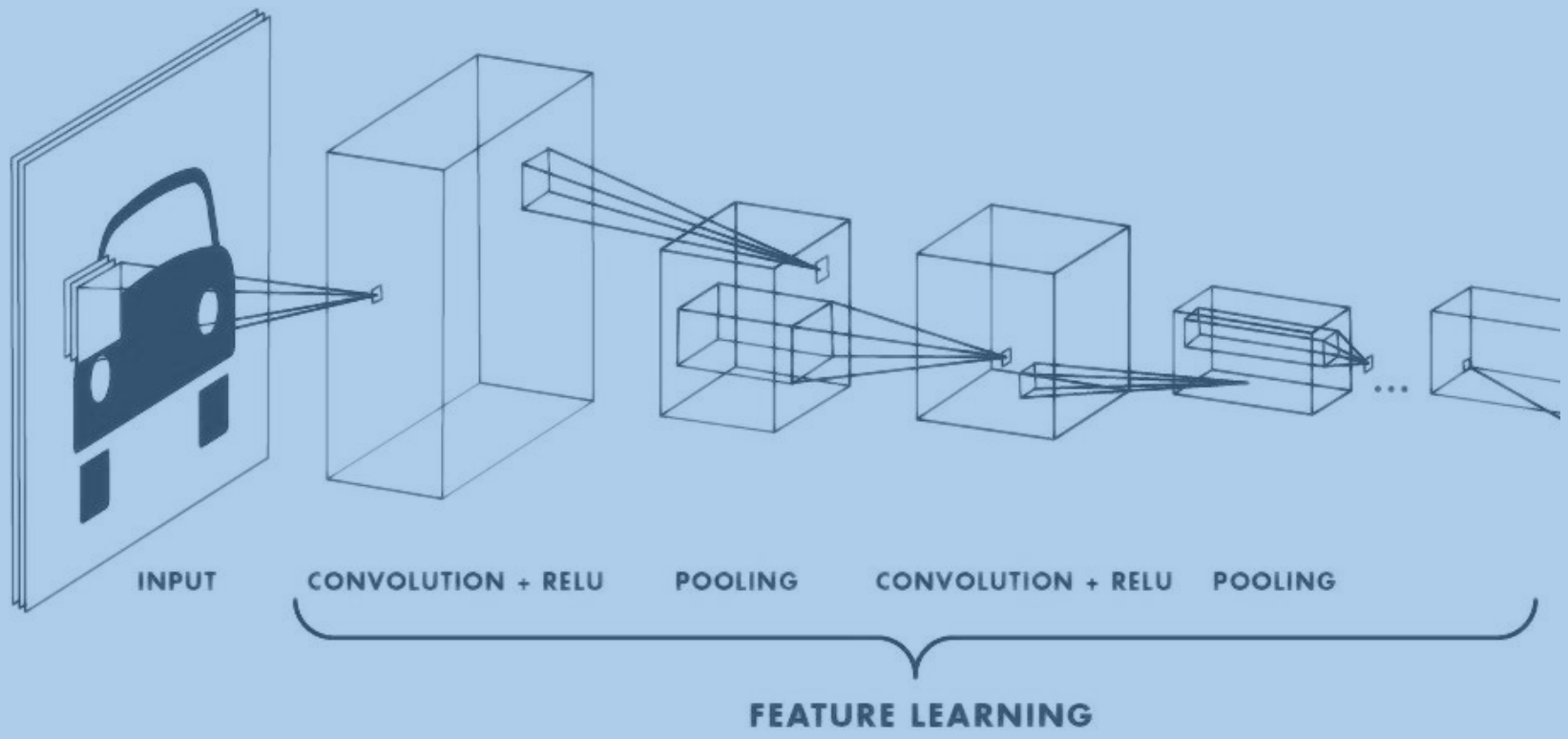
Deep CNN


INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FEATURE LEARNING

Deep CNN

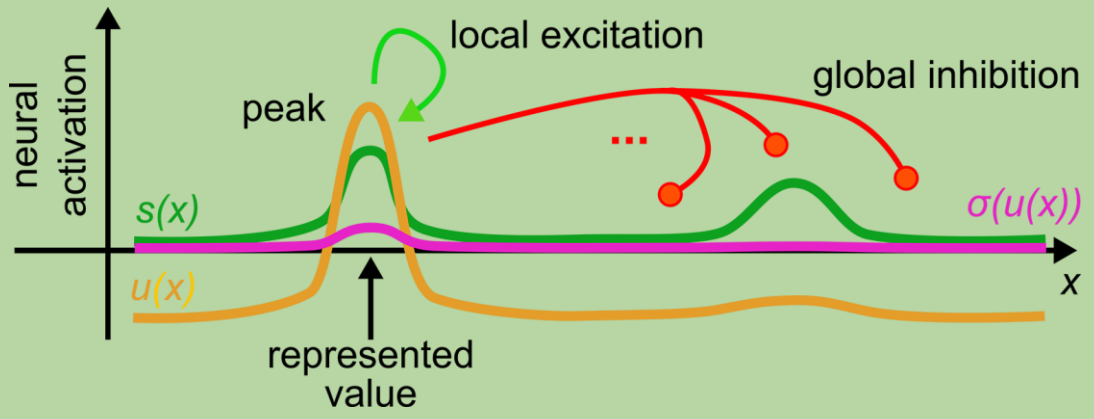

CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION

Deep CNN



Feature Maps

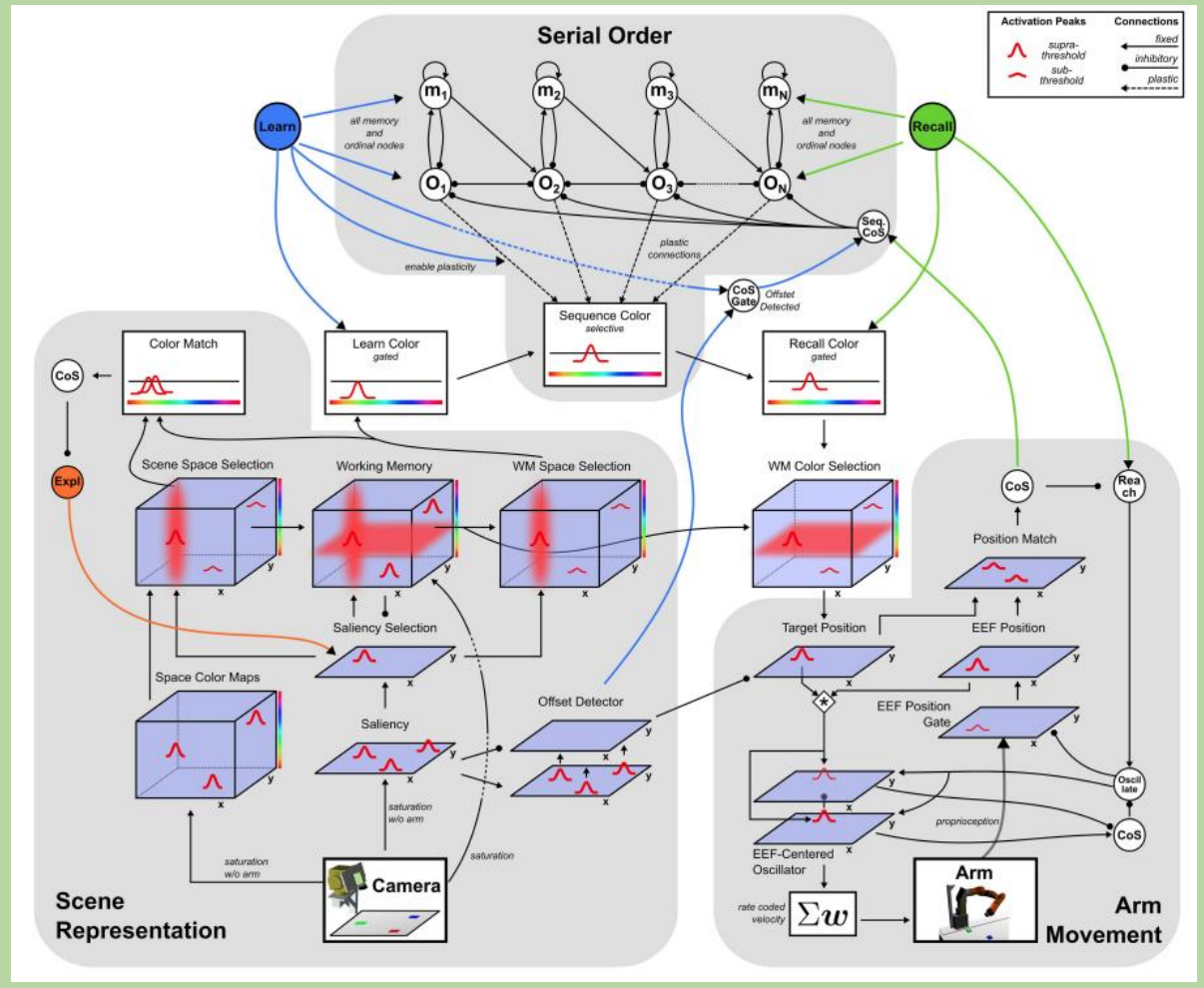
Dynamic Field Theory



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

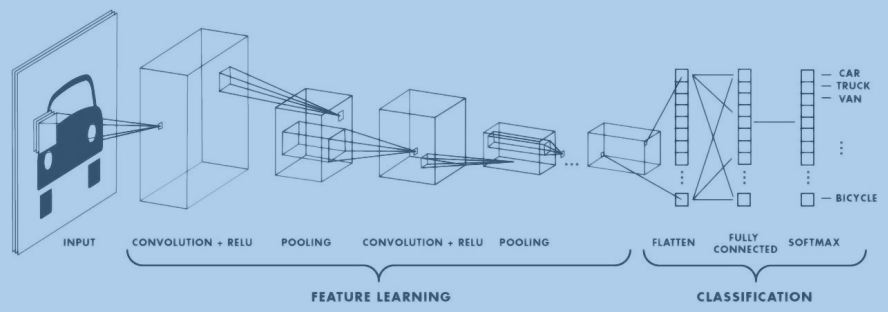
Schöner, G., Spencer, J. P., & The DFT Research Group (2016). *Dynamic thinking: A primer on dynamic field theory*. Oxford University Press

Dynamic Field Theory



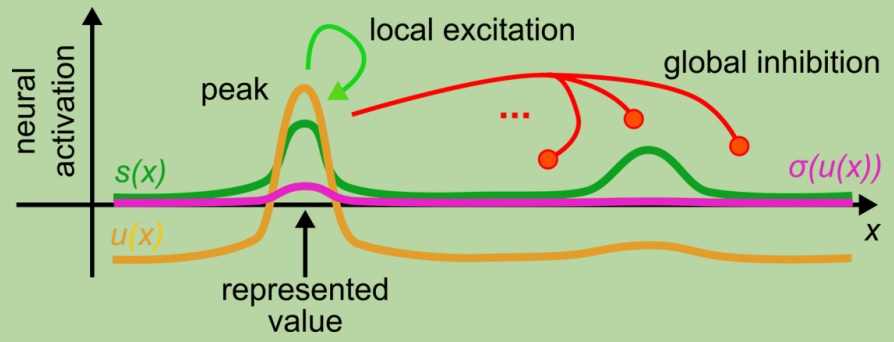
Tekülve, J., Fois, A., Sandamirskaya, Y., & Schöner, G.. (2019). Autonomous Sequence Generation for a Neural Dynamic Robot. Frontiers in Neurorobotics

Deep CNN



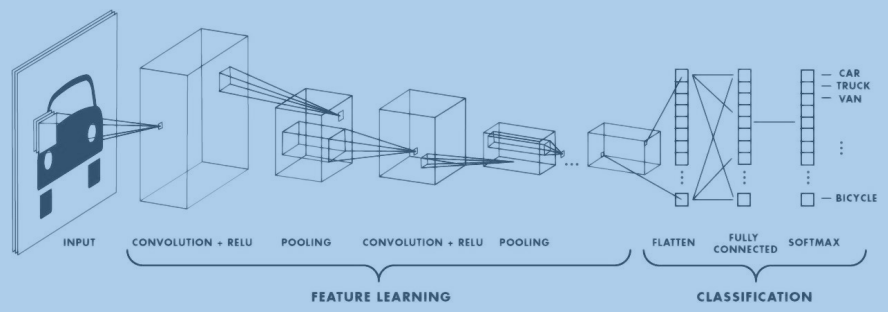
Feed-forward path

Dynamic Field Theory



Higher cognition

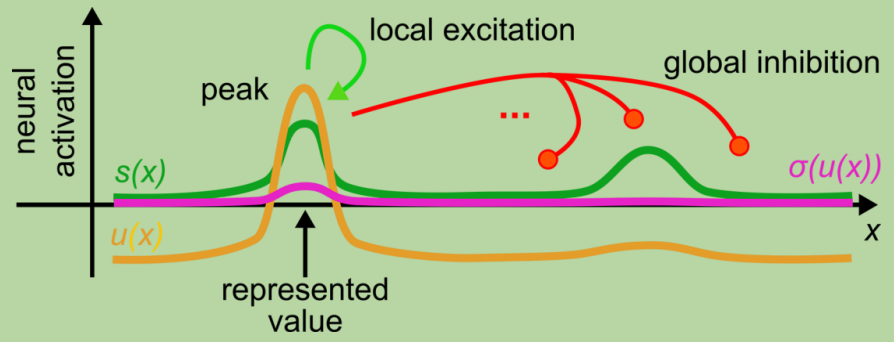
Deep CNN



Feed-forward path

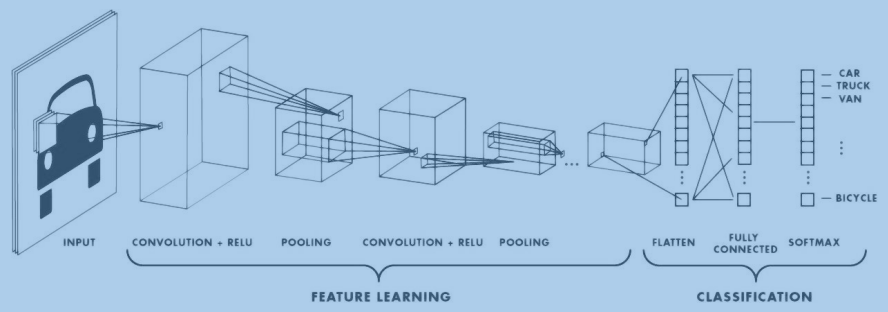
interface

Dynamic Field Theory



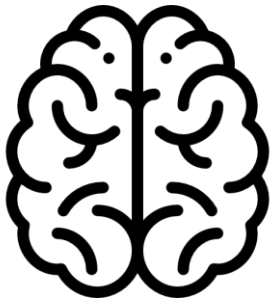
Higher cognition

Deep CNN

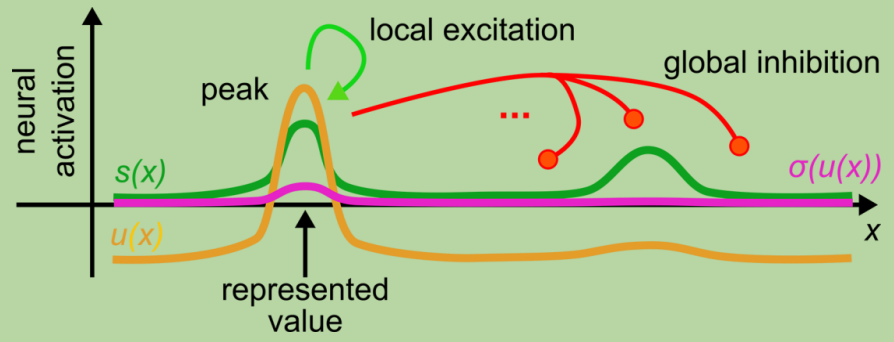


Feed-forward path

interface

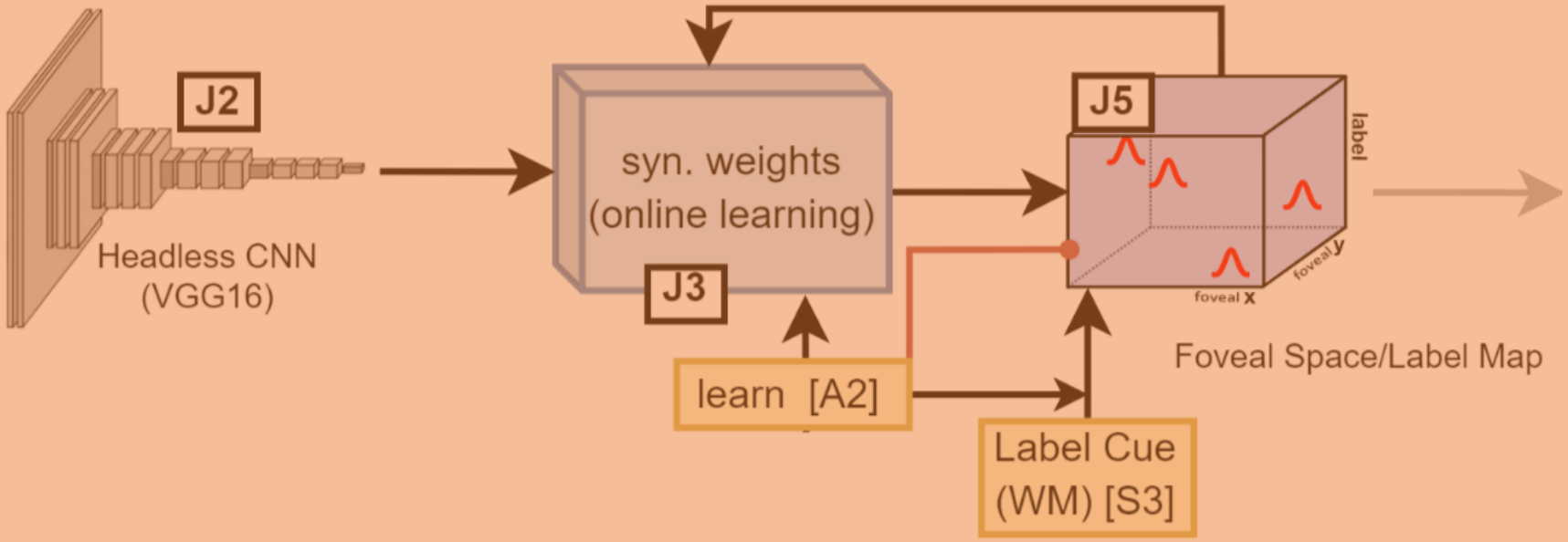


Dynamic Field Theory



Higher cognition

Interface

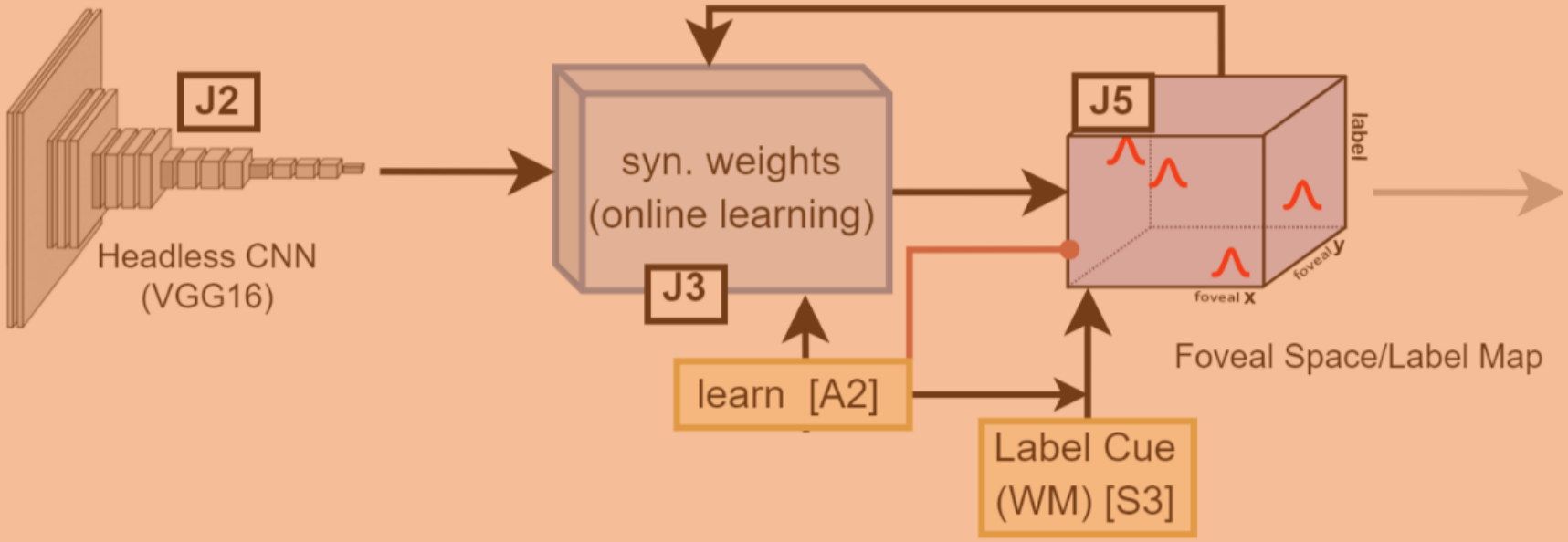


$$\tau_w \dot{w}_{m_\mathrm{f},u_\mathrm{fsml}}(\boldsymbol{x},t) = \eta\ \sigma(u_\mathrm{learn})\ y\ (y-\Theta)\ \frac{m_\mathrm{f}(x_1,x_2,t)}{\Theta}$$

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

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

Interface



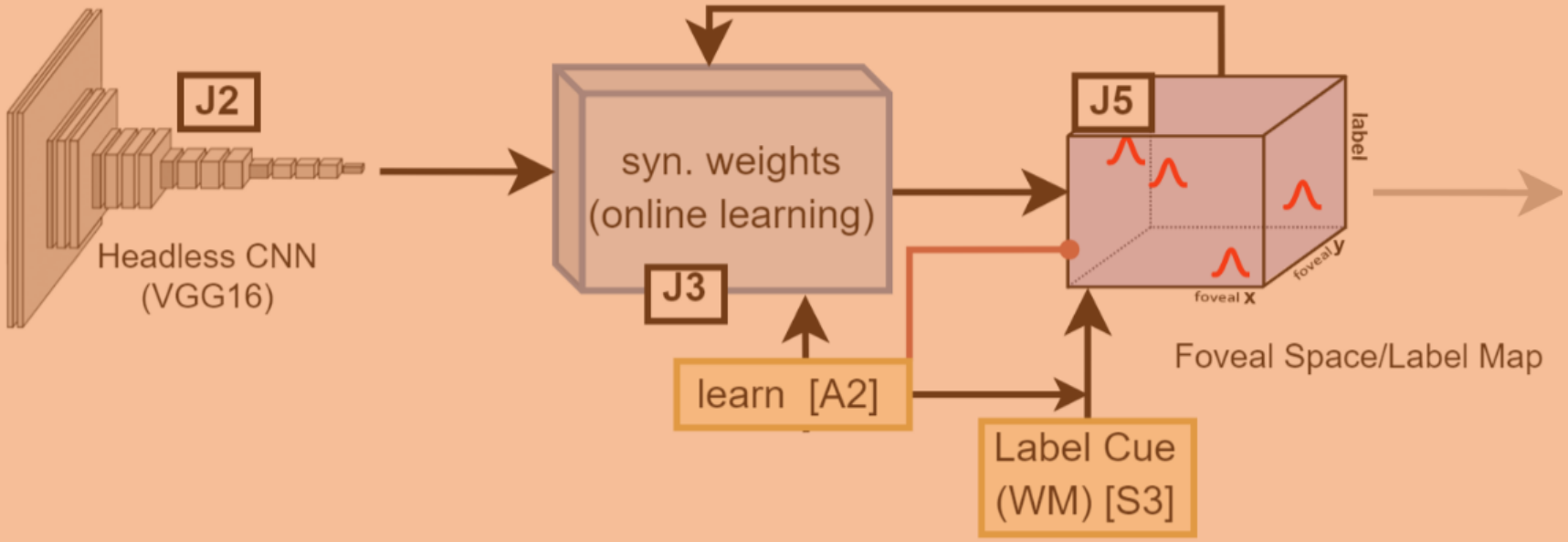
$$\tau_w \dot{w}_{m_f,u_{fsml}}(\boldsymbol{x},t) = \eta\ \sigma(u_{learn})\ y\ (y-\Theta)\ \frac{m_f(x_1,x_2,t)}{\Theta}$$

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

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

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

Interface

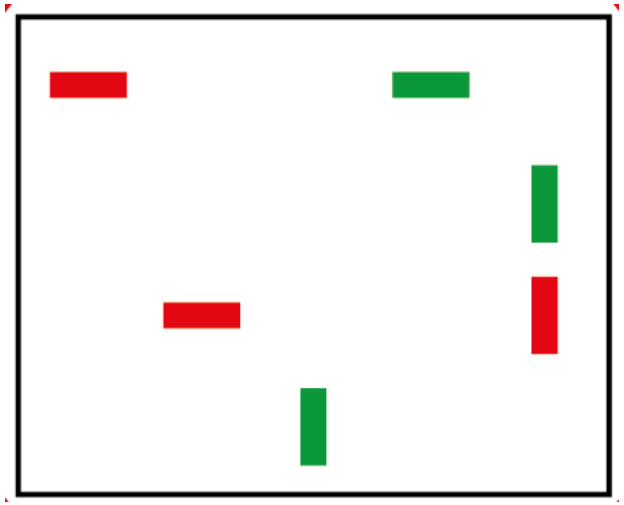


$$\tau_w \dot{w}_{m_\mathrm{f},u_\mathrm{fsml}}(\boldsymbol{x},t) = \eta\ \sigma(u_\mathrm{learn})\ y\ (y-\Theta)\ \frac{m_\mathrm{f}(x_1,x_2,t)}{\Theta}$$

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

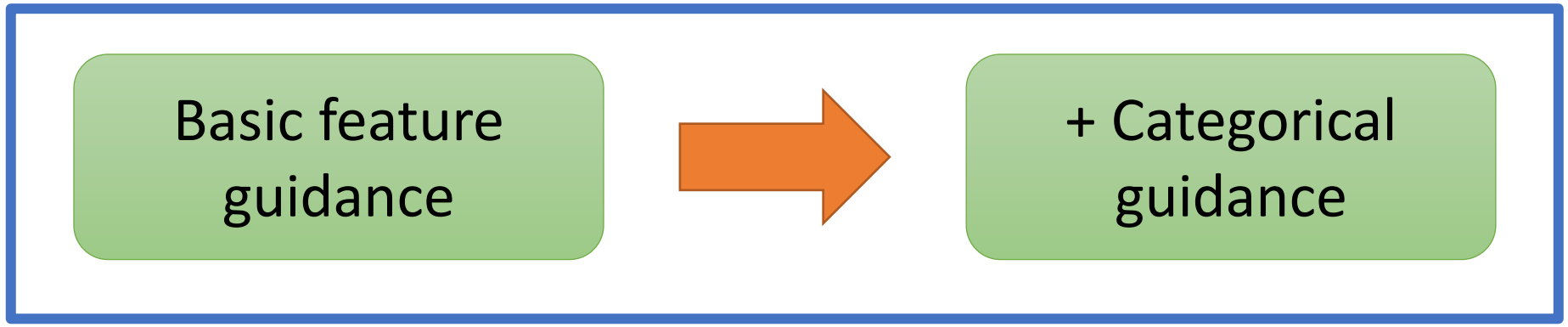
$$\tau_\Theta \dot{\Theta} = (y^2 - \Theta),$$

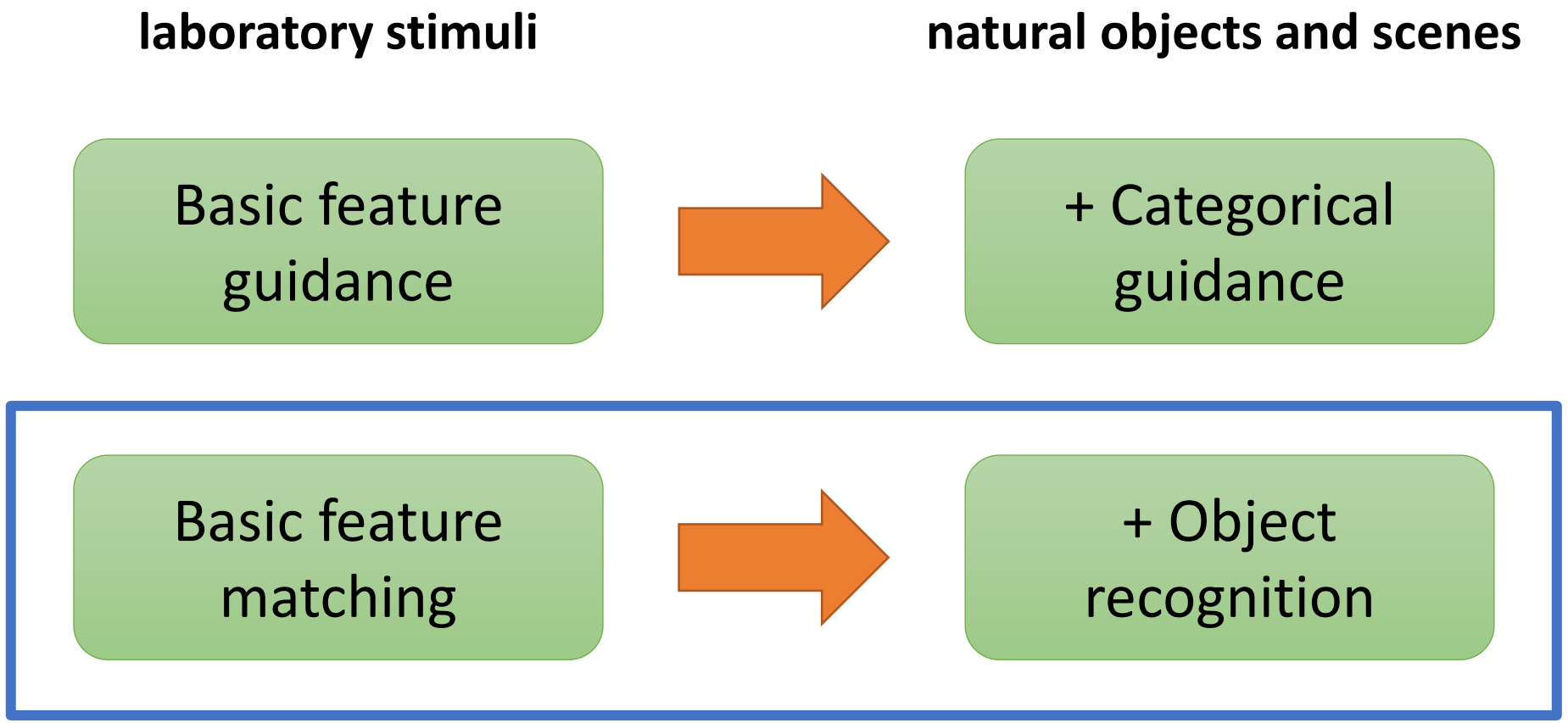
Lim, S., et al. (2015). Inferring learning rules from distributions of firing rates in cortical neurons. Nature neuroscience



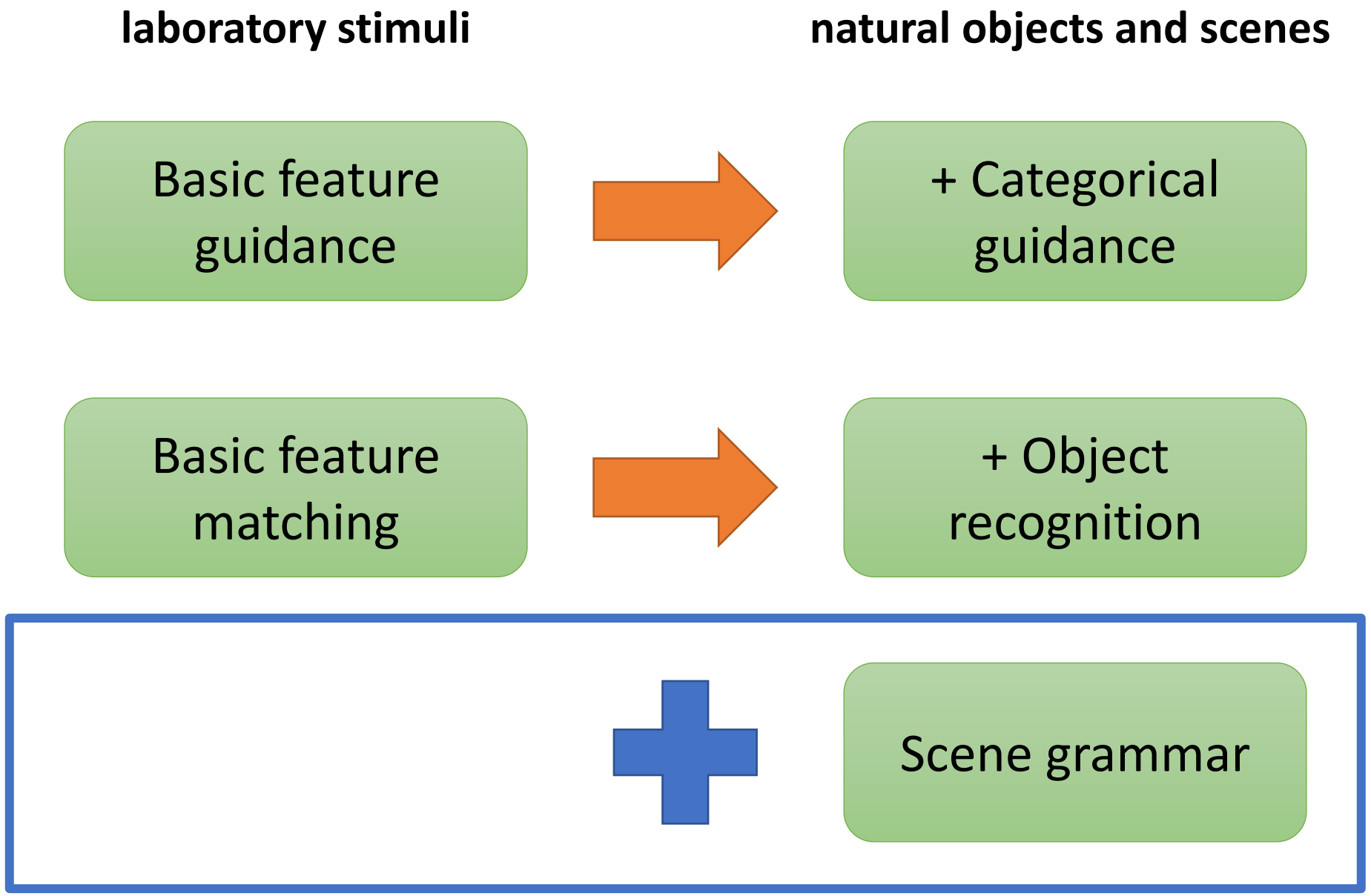

PRIVAT

laboratory stimuli **natural objects and scenes**

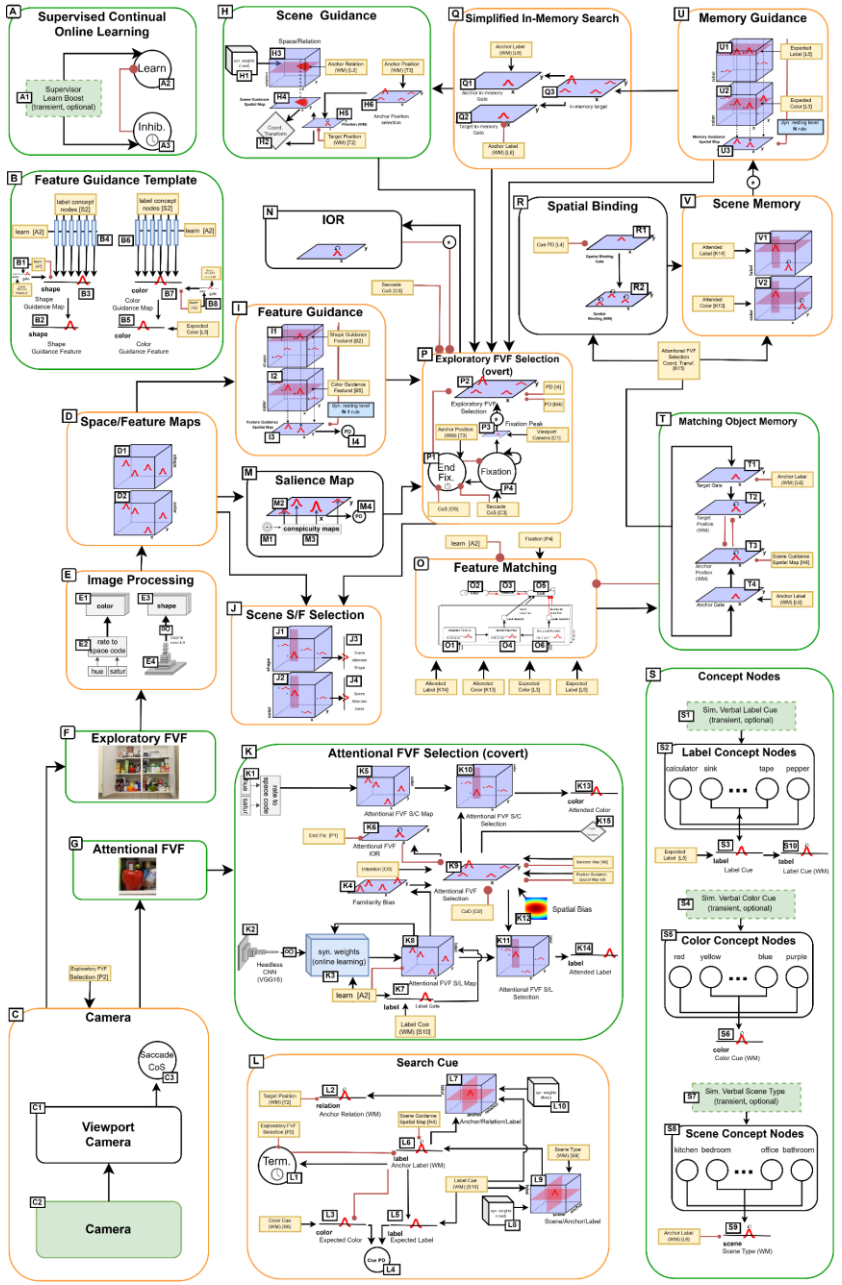



laboratory stimuli
natural objects and scenes
Basic feature guidance
+ Categorical guidance
Basic feature matching
+ Object recognition

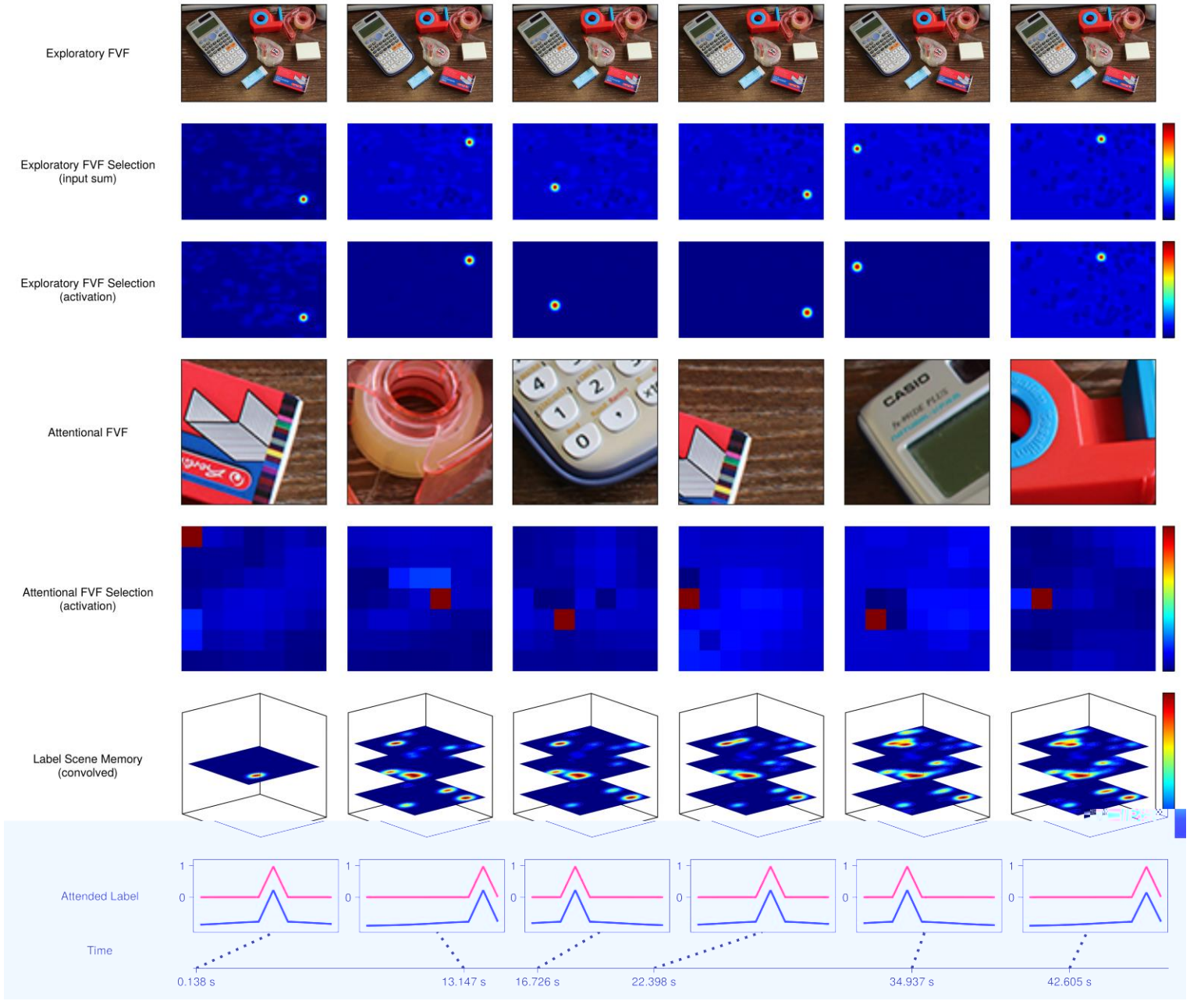
Boettcher, S. E., Draschkow, D., Dienhart, E., & Vo, M. L.- H. (2018). Anchoring visual search in scenes. Journal of vision


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

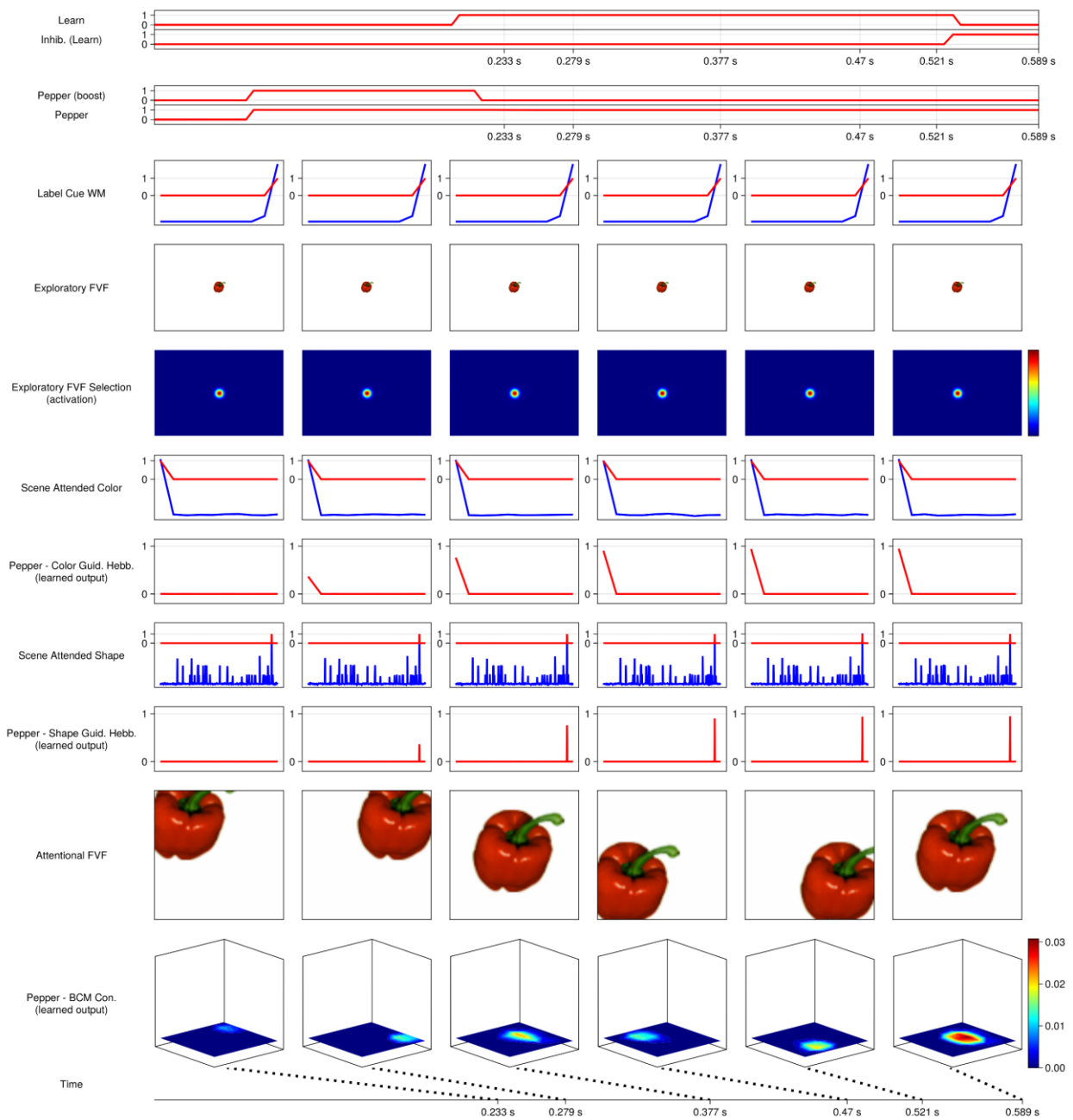
The neural process model



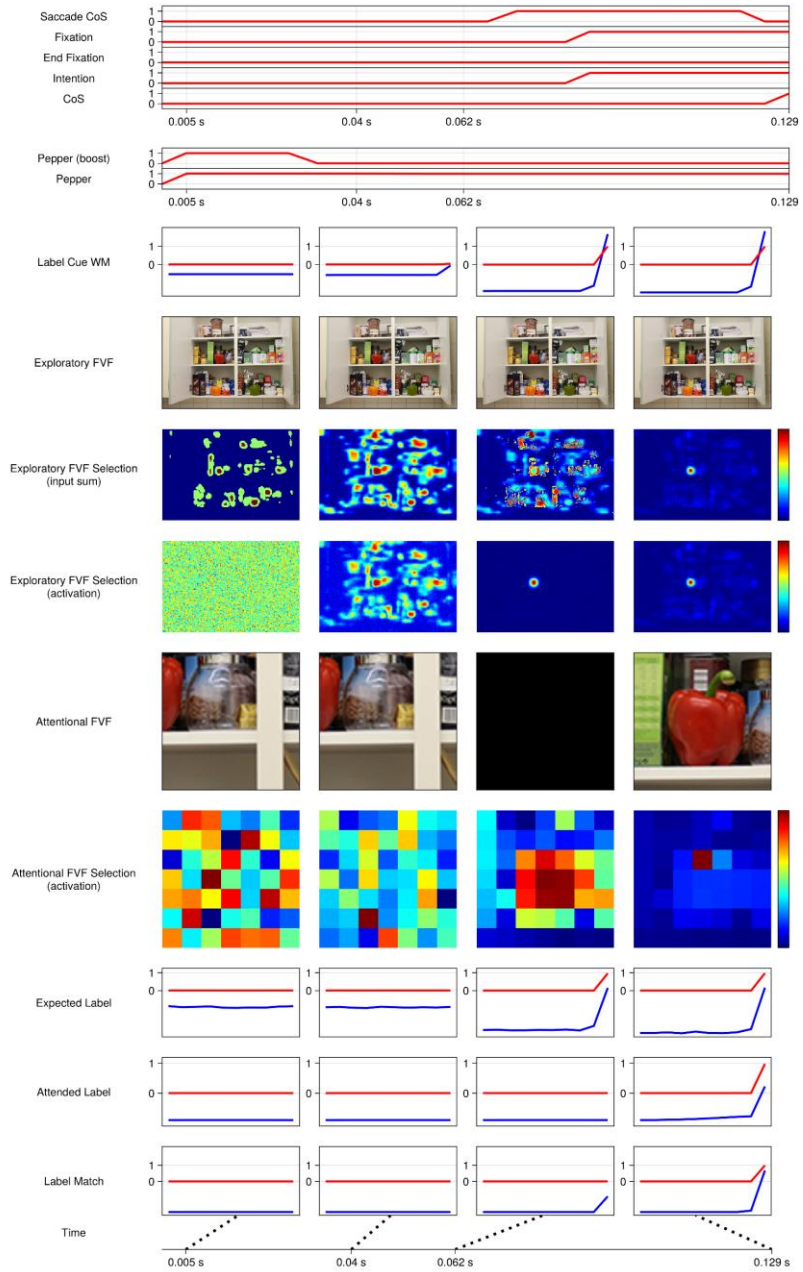
Results


Exploratory FVF
Exploratory FVF Selection (input sum)
Exploratory FVF Selection (activation)
Attentional FVF
Attentional FVF Selection (activation)
Label Scene Memory (convolved)
Attended Label
Time
1
0
0.138 s
13.147 s
16.726 s
22.398 s
34.937 s
42.605 s

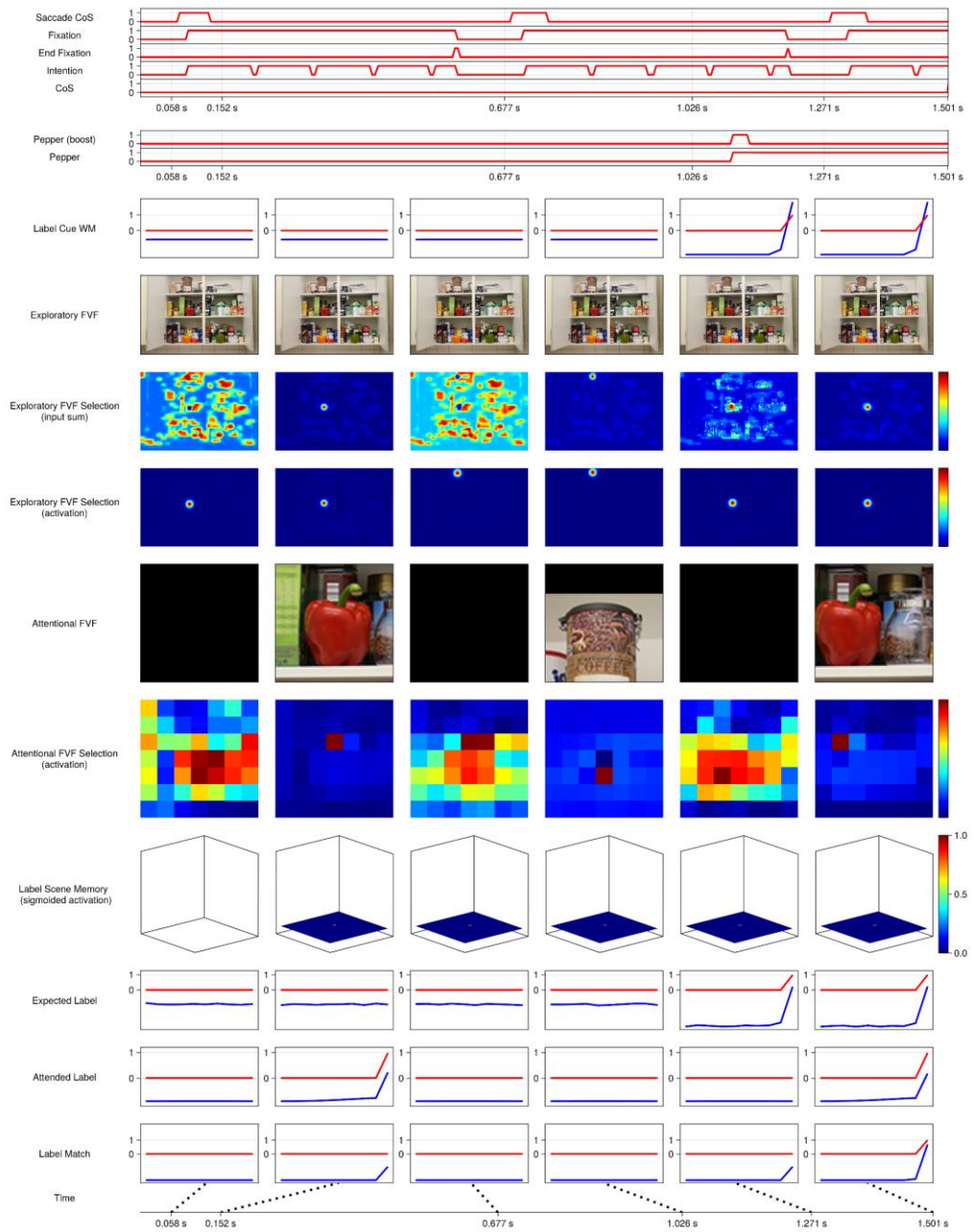
Results


Learn
Inhib. (Learn)
Pepper (boost)
Pepper
Label Cue WM
Exploratory FVF
Exploratory FVF Selection (activation)
Scene Attended Color
Pepper - Color Guid. Hebb. (learned output)
Scene Attended Shape
Pepper - Shape Guid. Hebb. (learned output)
Attentional FVF
Pepper - BCM Con. (learned output)
Time
0.233 s
0.279 s
0.377 s
0.47 s
0.521 s
0.589 s
0.03
0.02
0.01
0.00

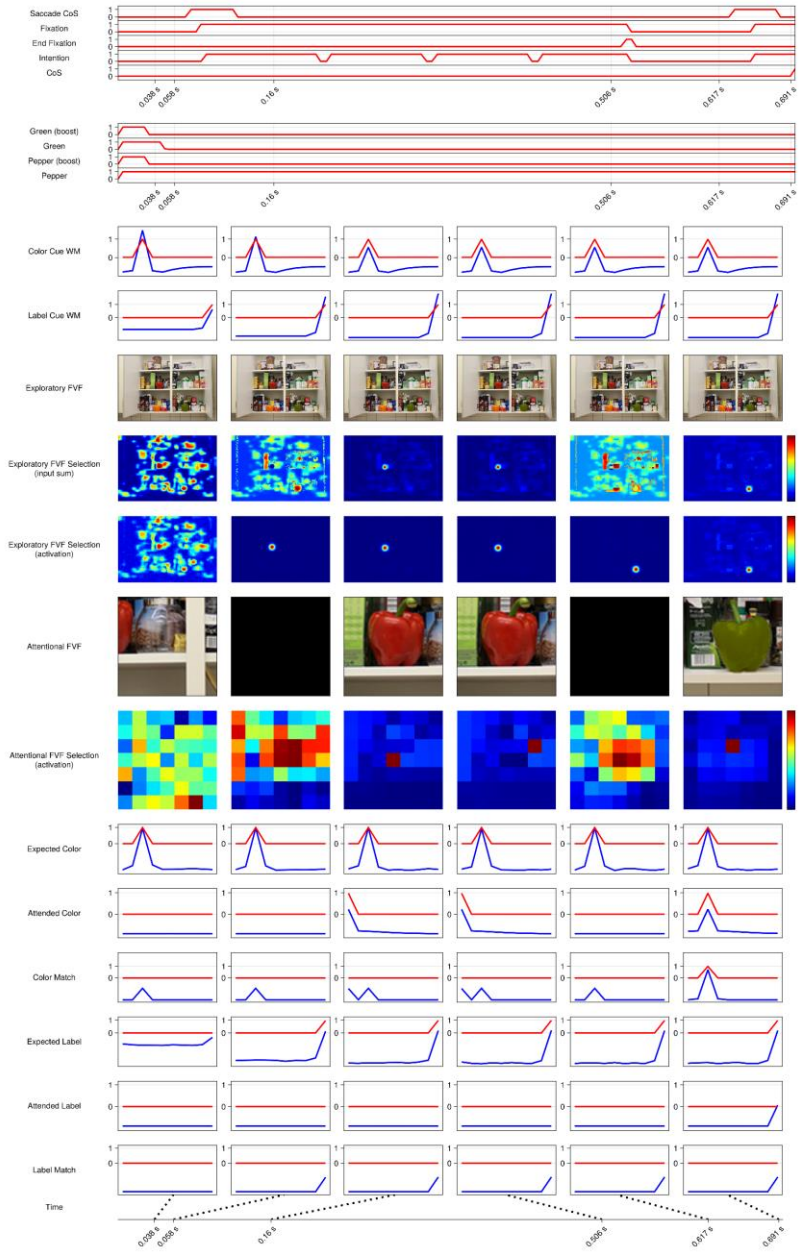
Results


Saccade CoS
Fixation
End Fixation
Intention
CoS
0.005 s
0.04 s
0.062 s
0.129
Pepper (boost)
Pepper
Label Cue WM
Exploratory FVF
Exploratory FVF Selection (input sum)
Exploratory FVF Selection (activation)
Attentional FVF
Attentional FVF Selection (activation)
Expected Label
Attended Label
Label Match
Time
0.129 s

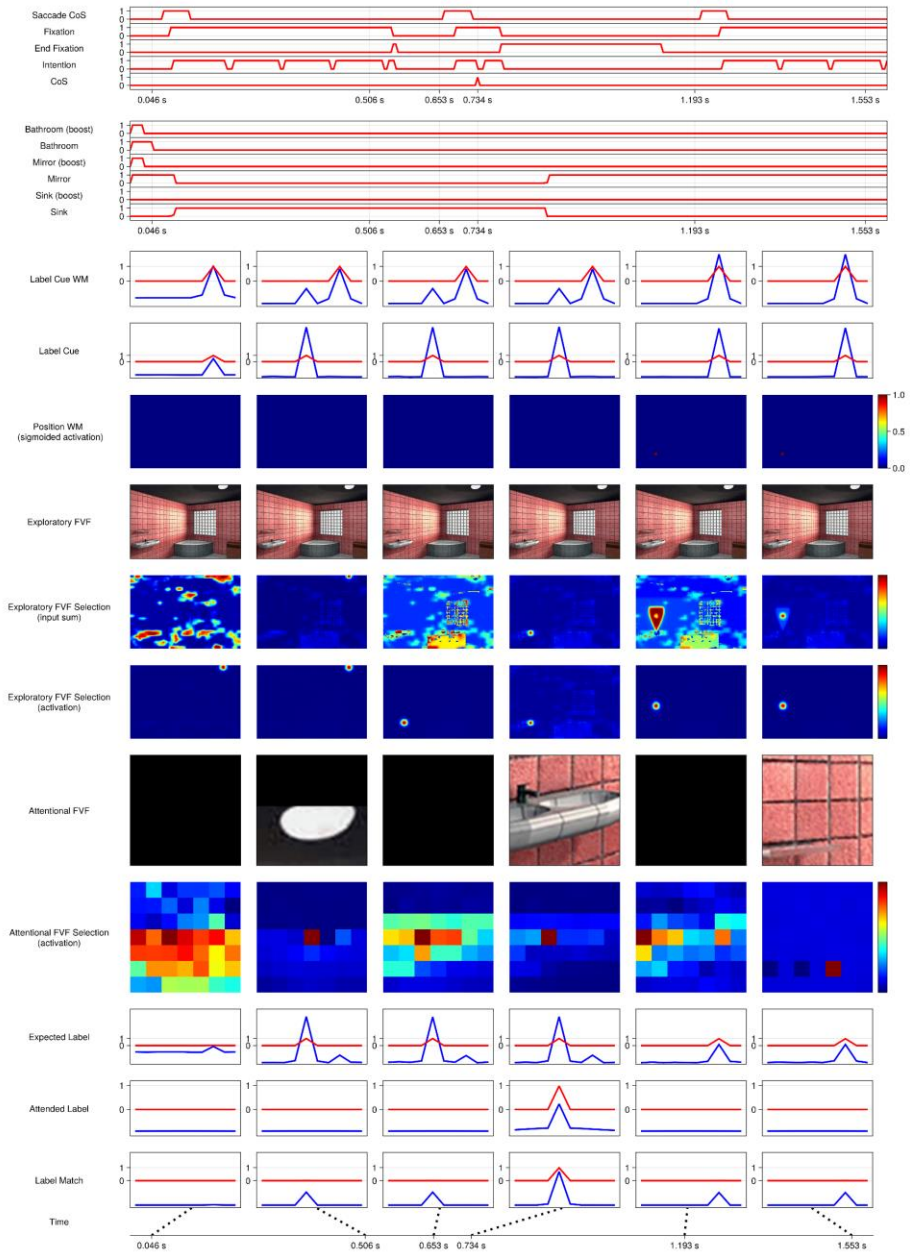
Results


Saccade CoS
Fixation
End Fixation
Intention
CoS
0.058 s
0.152 s
0.677 s
1.026 s
1.271 s
1.501 s
Pepper (boost)
Pepper
Label Cue WM
Exploratory FVF
Exploratory FVF Selection (input sum)
Exploratory FVF Selection (activation)
Attentional FVF
Attentional FVF Selection (activation)
Label Scene Memory (sigmoided activation)
1.0
0.5
0.0
Expected Label
Attended Label
Label Match
Time

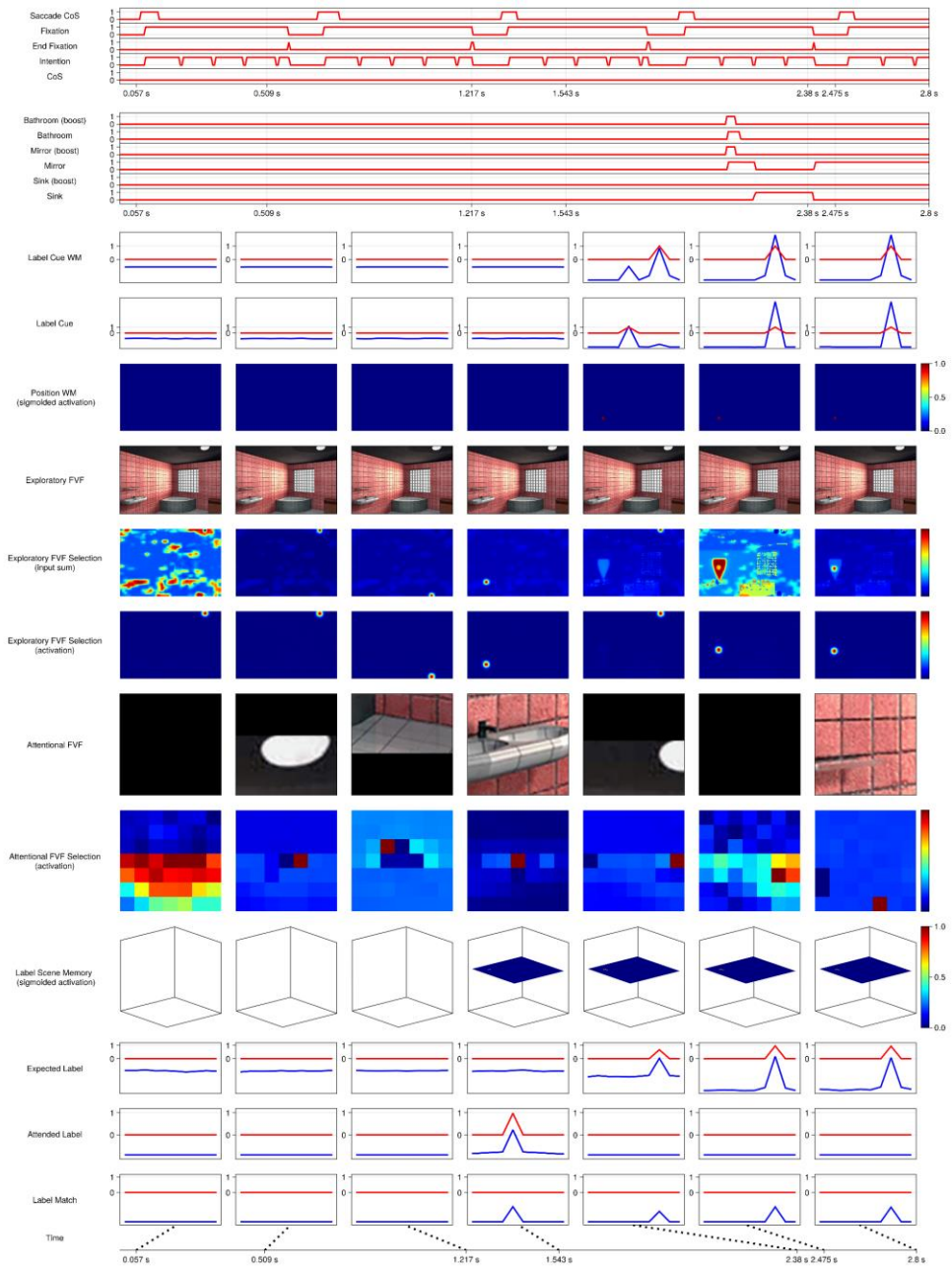
Results


Saccade CoS
Fixation
End Fixation
Intention
CoS
Green (boost)
Green
Pepper (boost)
Pepper
Color Cue WM
Label Cue WM
Exploratory FVF
Exploratory FVF Selection (input sum)
Exploratory FVF Selection (activation)
Attentional FVF
Attentional FVF Selection (activation)
Expected Color
Attended Color
Color Match
Expected Label
Attended Label
Label Match
Time
0.038 s
0.058 s
0.16 s
0.506 s
0.617 s
0.691 s

Results


Saccade CoS
Fixation
End Fixation
Intention
CoS
0.046 s
0.506 s
0.653 s
0.734 s
1.193 s
1.553 s
Bathroom (boost)
Bathroom
Mirror (boost)
Mirror
Sink (boost)
Sink
Label Cue WM
Label Cue
Position WM (sigmoided activation)
1.0
0.5
0.0
Exploratory FVF
Exploratory FVF Selection (input sum)
Exploratory FVF Selection (activation)
Attentional FVF
Attentional FVF Selection (activation)
Expected Label
Attended Label
Label Match
Time

Results


Saccade CoS
Fixation
End Fixation
Intention
CoS
0.057 s
0.509 s
1.217 s
1.543 s
2.38 s 2.475 s
2.8 s
Bathroom (boost)
Bathroom
Mirror (boost)
Mirror
Sink (boost)
Sink
Label Cue WM
Label Cue
Position WM (sigmoided activation)
Exploratory FVF
Exploratory FVF Selection (input sum)
Exploratory FVF Selection (activation)
Attentional FVF
Attentional FVF Selection (activation)
Label Scene Memory (sigmoided activation)
Expected Label
Attended Label
Label Match
Time

- 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