HMAX Model

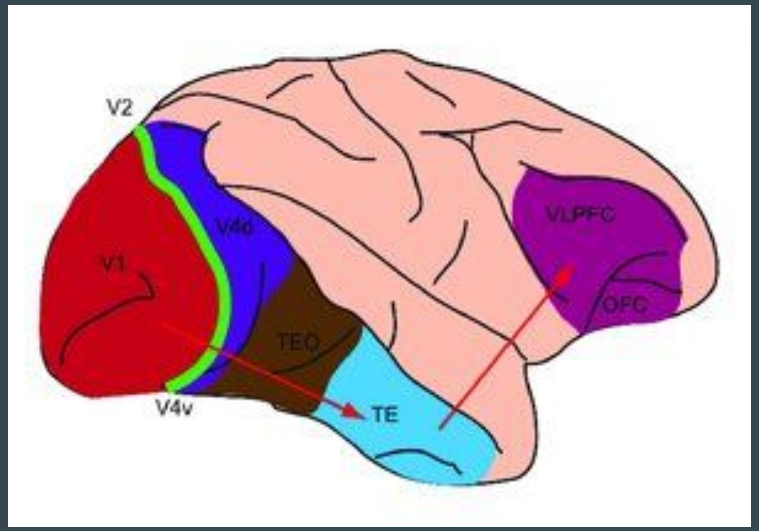
Jenish Maharjan
Rahul Thapa

Introduction

- Computational model of object recognition in Cortex
- Riesenhuber & Poggio, Nature Neuroscience, 1999
- Serre & Riesenhuber, 2004
- Dubbed by Mike Tarr as “Hierarchical Model and X”
- Summarizes the basic facts about the ventral visual stream, a hierarchy of brain areas thought to mediate object recognition in cortex

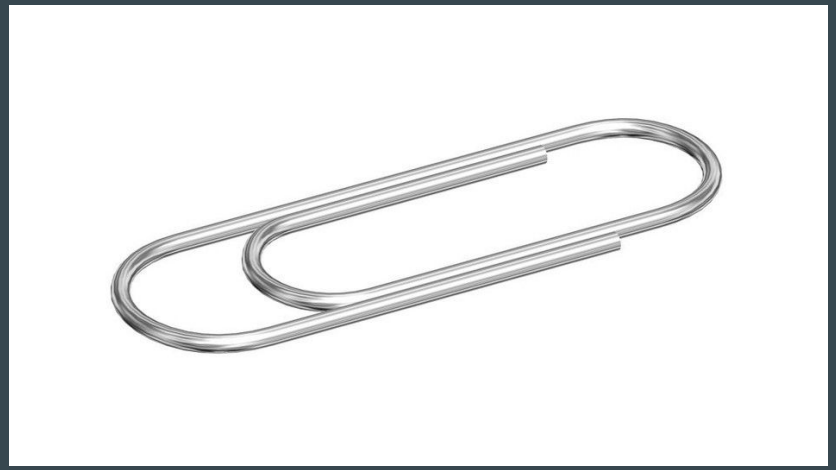
Ventral Visual Pathway

- Mediates Object recognition in cortex
- V1 - V2 - V4 - IT → PFC
- Cells found in macaque IT - key role in object recognition
- Hallmark of those cells - Robustness of their responses to stimulus transformations such as scale and position changes



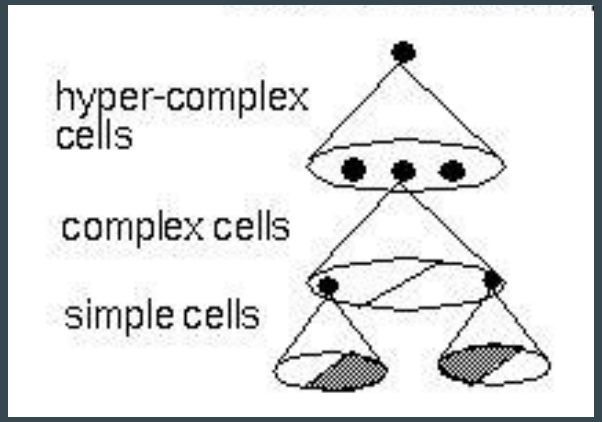
Key Goals of Object Recognition

- Invariance - the ability to recognize a pattern under various transformations
- Specificity - the ability to discriminate between different patterns.

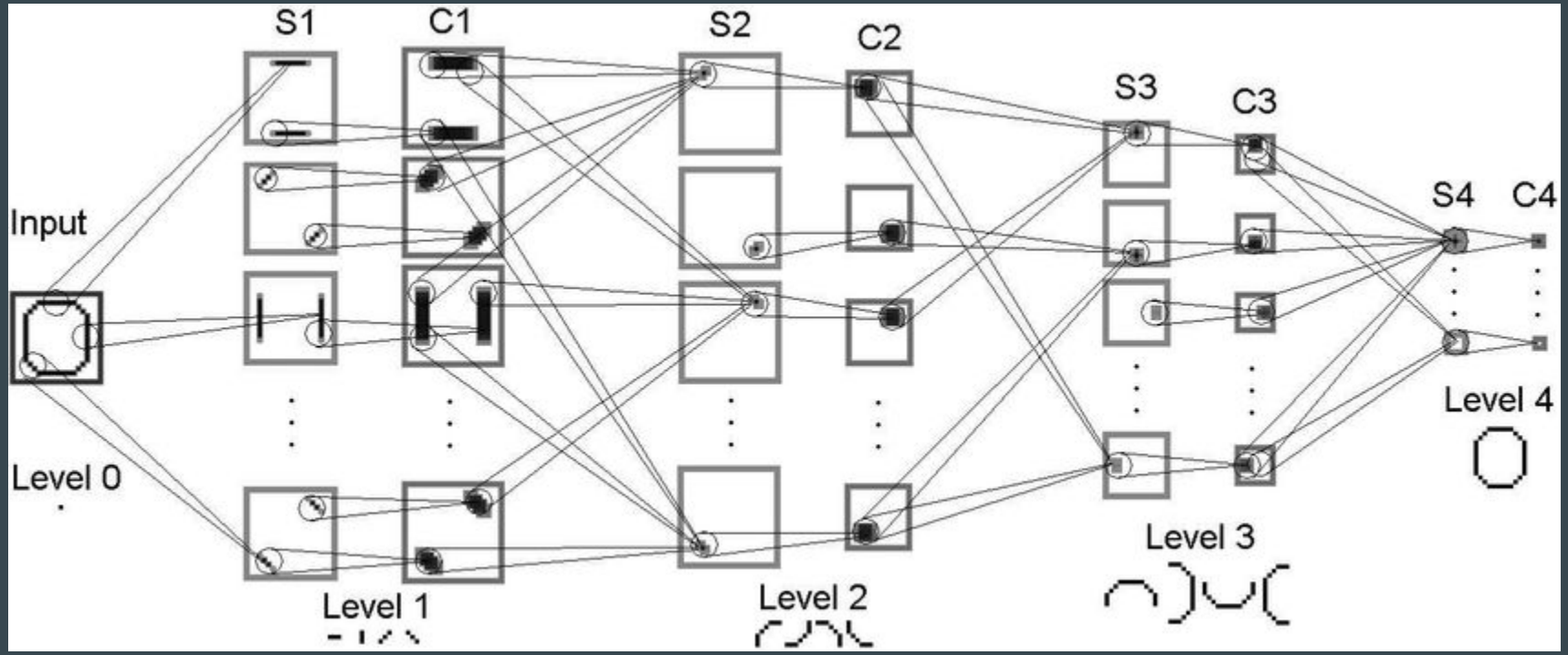


Hubel and Wiesel

- Model of Simple Cells and Complex Cells
- Simple Cells - Small Receptive Fields with phase dependence
- Complex Cells - Large Receptive Fields with no phase dependence
- Simple cells with neighboring receptive fields feed into the same complex cell, thereby endowing that complex cell with a phase-invariant response



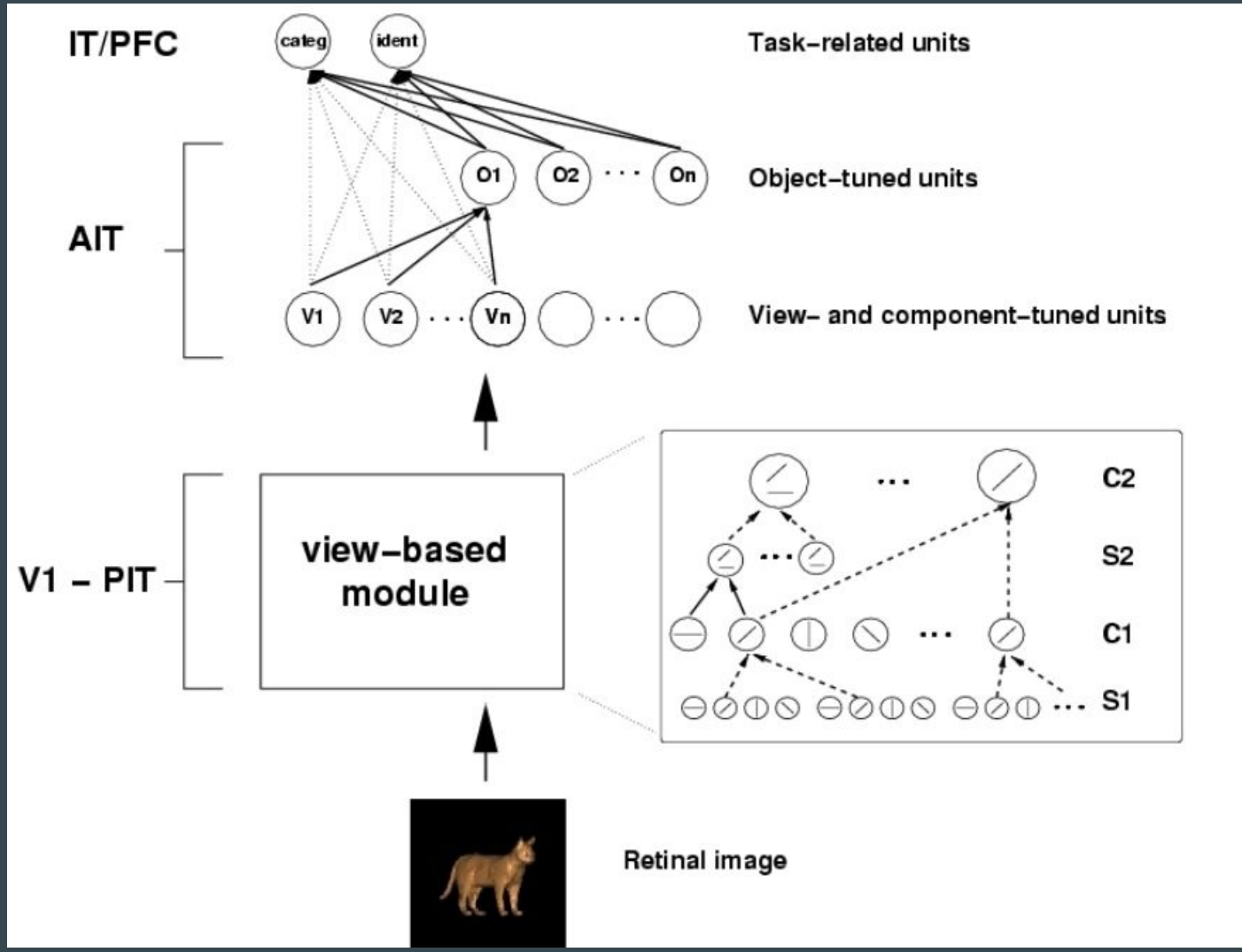
Fukushima - Neocognitron



Summary of Ideas behind the HMAX model

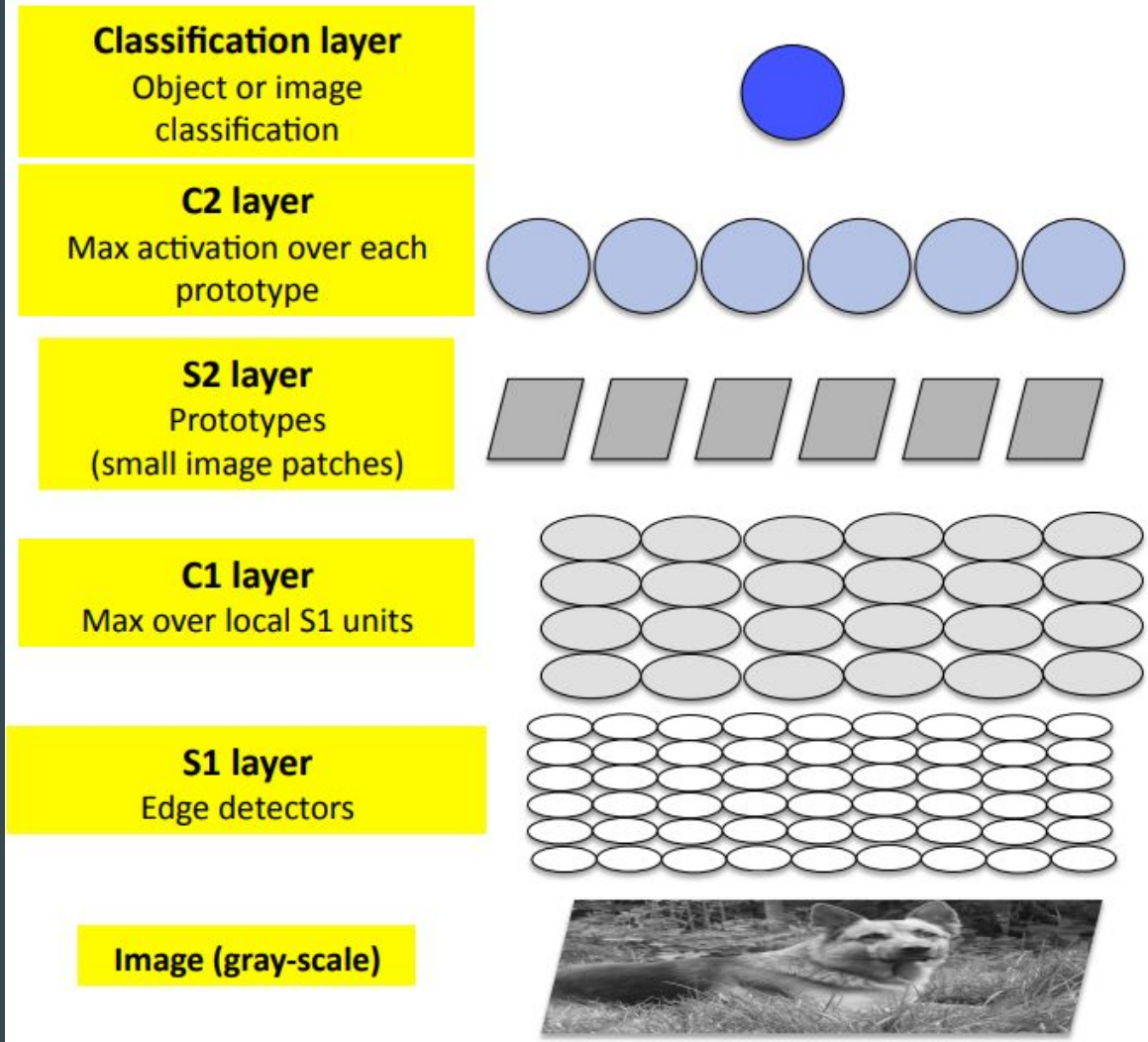
- “Immediate” visual processing is feedforward and hierarchical: low levels detect simple features, which are combined hierarchically into increasingly complex features to be detected
- Layers of hierarchy alternate between “sensitivity” (to detecting features) and “invariance” (to position, scale, orientation)
- Size of receptive fields increases along the hierarchy
- Degree of invariance increases along the hierarchy

HMax Model


IT/PFC
categ
ident
Task–related units
AIT
O1
O2
On
Object–tuned units
V1
V2
Vn
View– and component–tuned units
V1 – PIT
view–based module
C2
S2
C1
S1
Retinal image

Core of the HMAX Model

Task: To produce a higher level representation of an image that will be useful for classification.



Simple Cells S1

- Input images are densely sampled by arrays of two-dimensional filters Gx,y (second derivative of Gaussians) that can be expressed as:

$$G_{x,y} = \frac{(-x\cos\theta + y\sin\theta)^2}{\sigma^2(\sigma^2 - 1)}$$
$$\exp\left(-\frac{(x\cos\theta + y\sin\theta)^2 + (-x\cos\theta + y\sin\theta)^2}{2\sigma^2}\right)$$

- Output: -1 to 1, modeling simple cells of phase 0 and π

Complex Cells C1

- C1 unit’s activity is determined by the strongest input it receives
- This mechanisms rely on a non-linear MAX operation (or its soft-MAX approximation)
- SUM (linear Summation with Equal weights) vs MAX (a nonlinear maximum operation)

Simple Cells S2

- S2 unit response function is a Gaussian with mean 1 (i.e., {1, 1, 1, 1}) and standard deviation 1
- S2 unit has a maximal firing rate of 1 which is attained if each of its four afferents fires at a rate of 1 as well

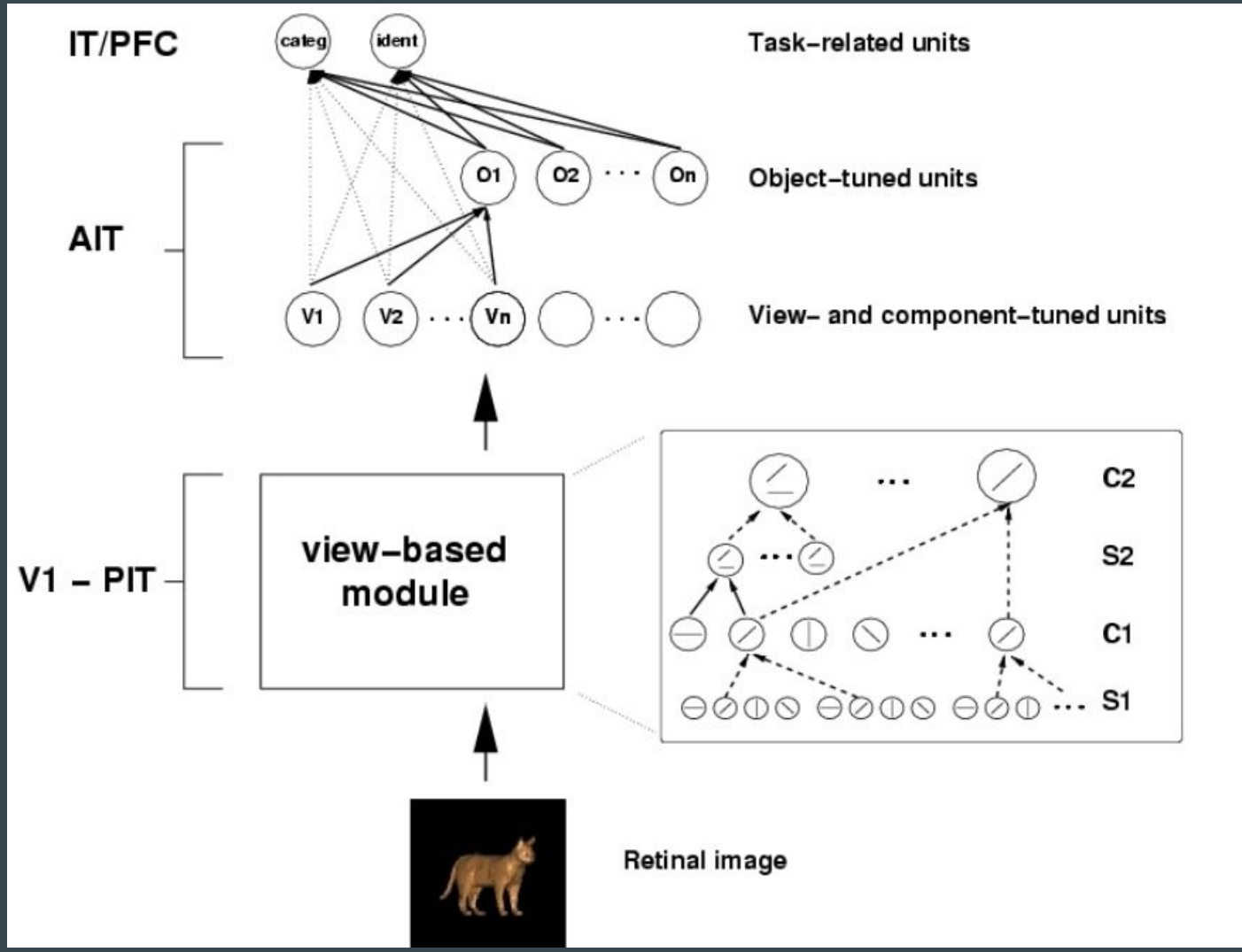
Complex Cells C2

- To finally achieve size invariance over all filter sizes in the four filter bands and position invariance over the whole input image, the S2 units are again pooled by a MAX operation to yield C2 units
- Output units of the HMAX core system, designed to correspond to neurons in extrastriate visual area V4 or posterior IT (PIT)

View Tuned Units

- C2 units in turn provide input to the view-tuned units (VTUs),
- VTU respond to a specific two-dimensional view of a three-dimensional object.
- The C2 → VTU connections are the only stage of the HMAX model where learning occurs.

HMax Model


IT/PFC
categ
ident
Task–related units
AIT
O1
O2
On
Object–tuned units
V1
V2
Vn
View– and component–tuned units
V1 – PIT
view–based module
C2
S2
C1
S1
Retinal image

Thank You!