Networks and Hierarchical Processing: Object Recognition in Human and Computer Vision

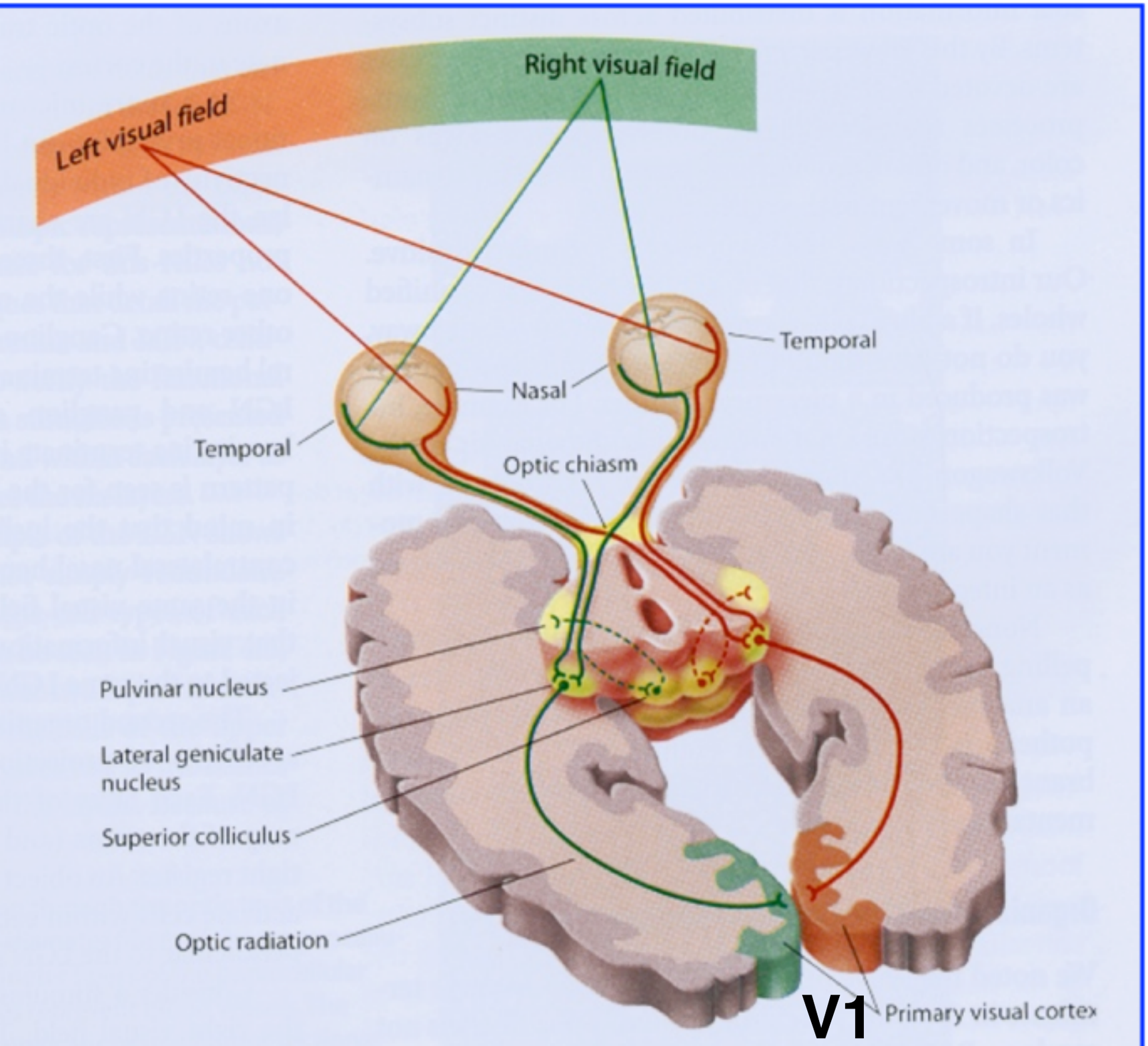
Guest Lecture: Marius Cătălin Iordan
CS 131 - Computer Vision: Foundations and Applications
01 December 2014

1. **Processing Pathways in the Human Visual System**
    - "what" and "where" pathways
    - building features in the ventral stream

2. **Hierarchical Pattern Recognition Systems**
    - early stages: small scale neural network
    - injecting neuroscience knowledge into design

3. **Third Age of the Neural Network: Modern Deep Nets**
    - extremely large scale (data + computation)
    - sudden, huge performance boost for recognition

From Retina to Cortex



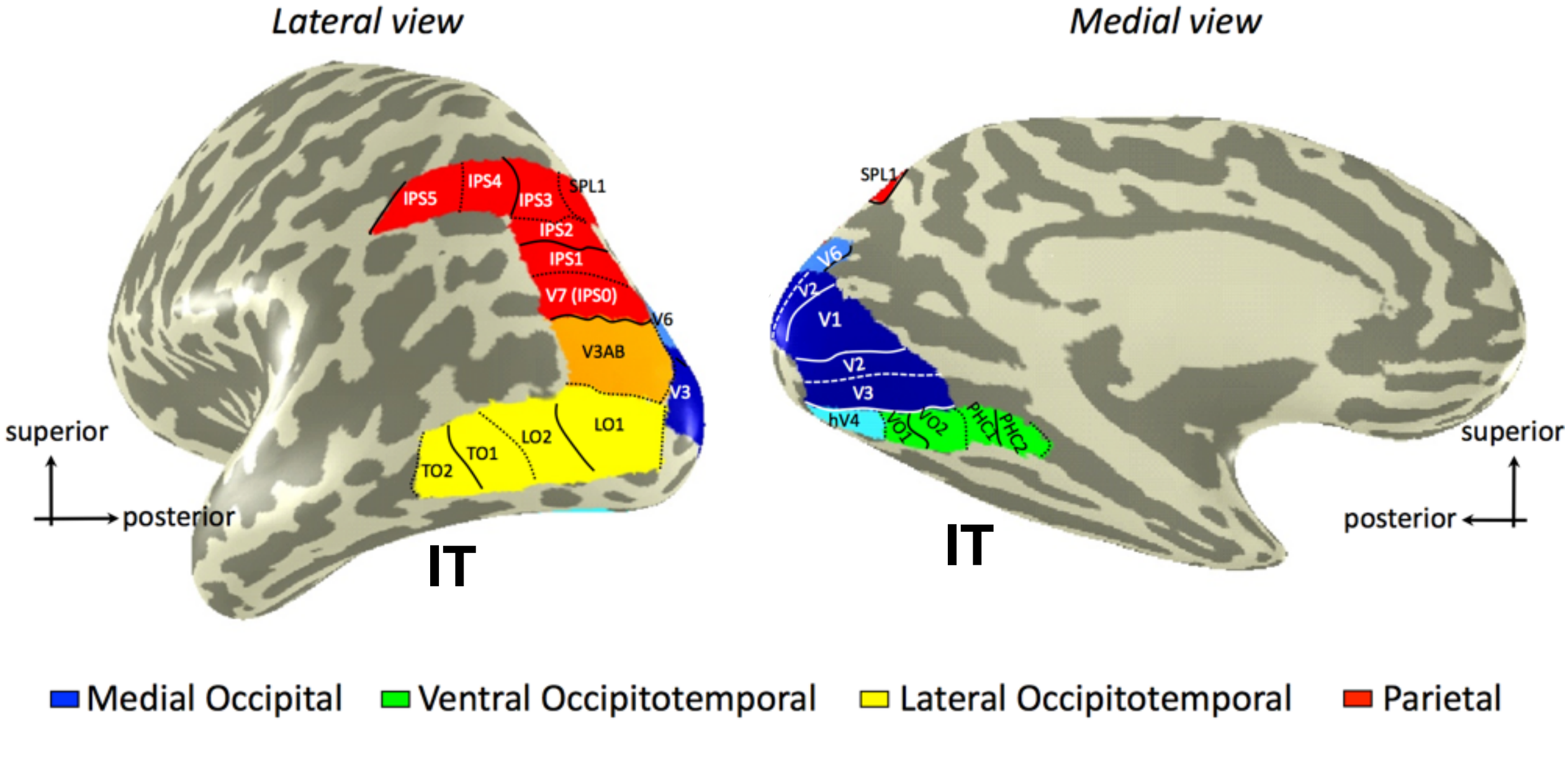
world

retina
(compression)

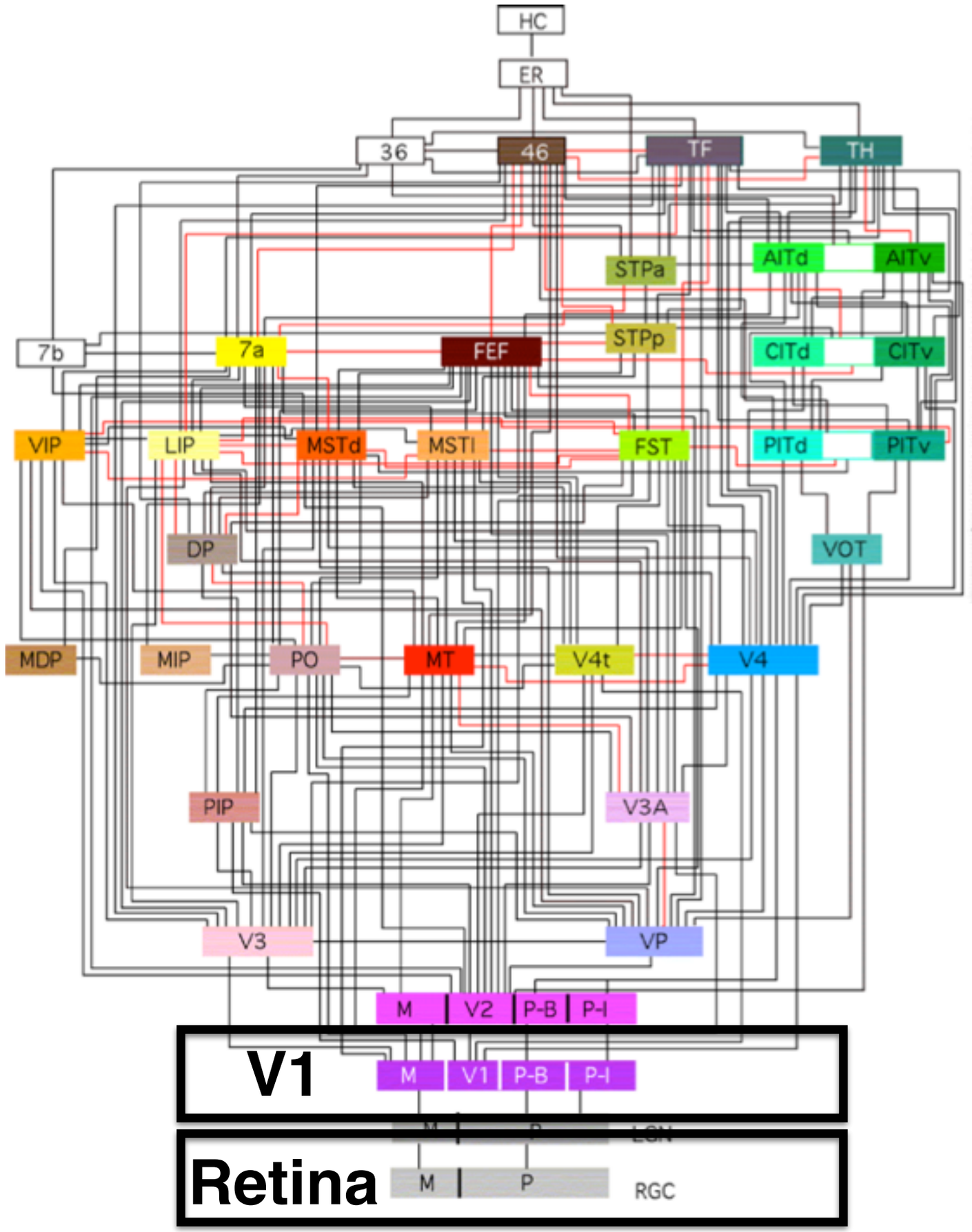
LGN

visual cortex
(expansion)

The Flow of Information



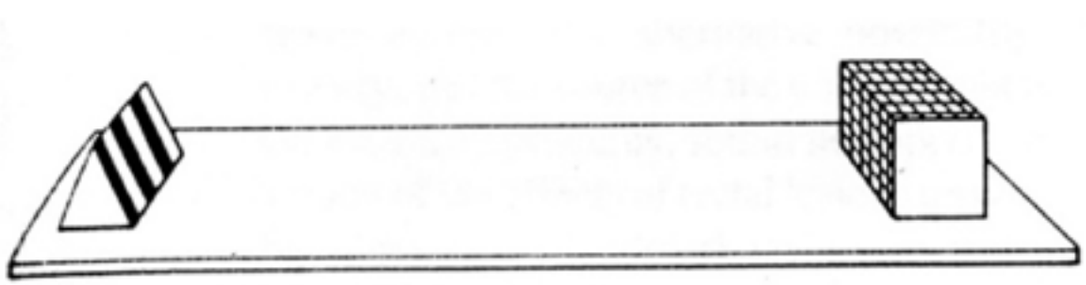
Weiner & Grill-Spector (2012)



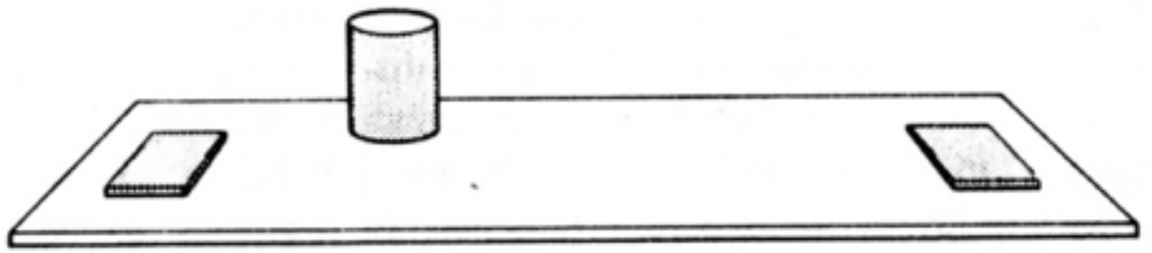
Van Essen (1991)

Specialization: “What” and “Where” Pathways

monkey lesion studies



“what”



“where”

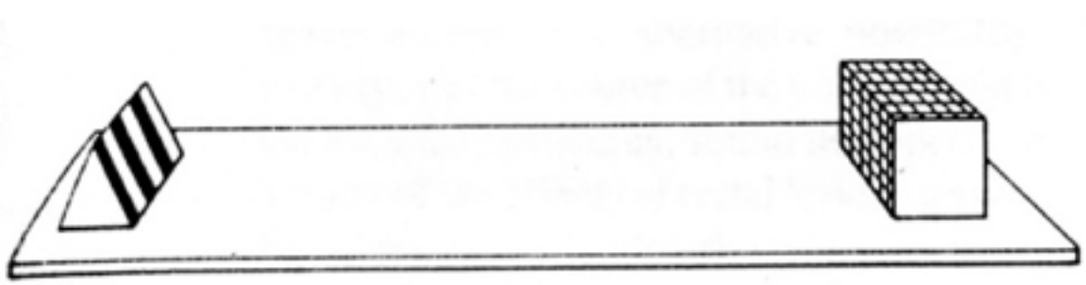
Mishkin & Ungerleider 1982

Specialization: “What” and “Where” Pathways

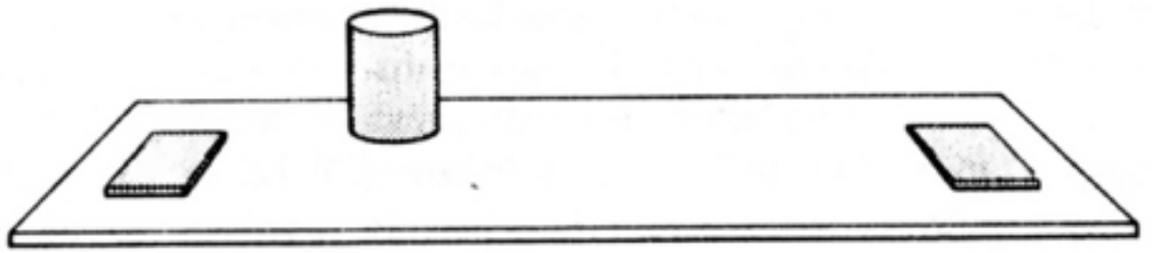
monkey lesion studies

lesion “where” pathway: difficulty in spatial reasoning

lesion “what” pathway: difficulty in object recognition



“what”



“where”

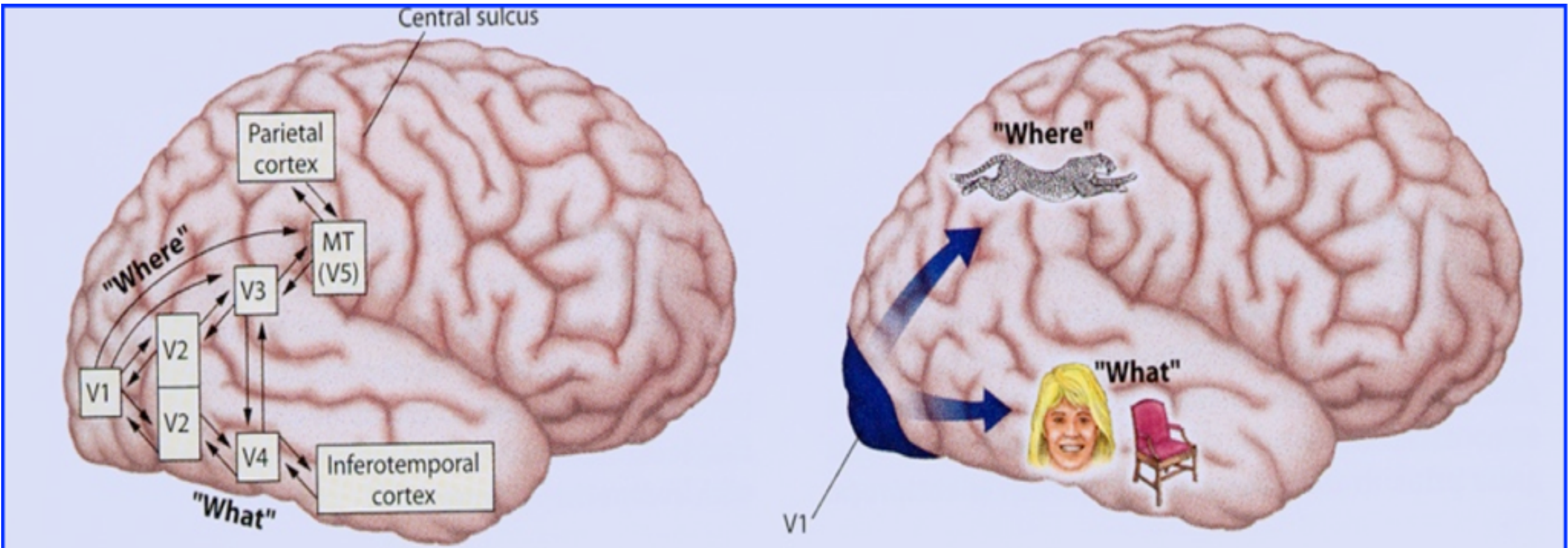
Mishkin & Ungerleider 1982

Specialization: “What” and “Where” Pathways

monkey lesion studies

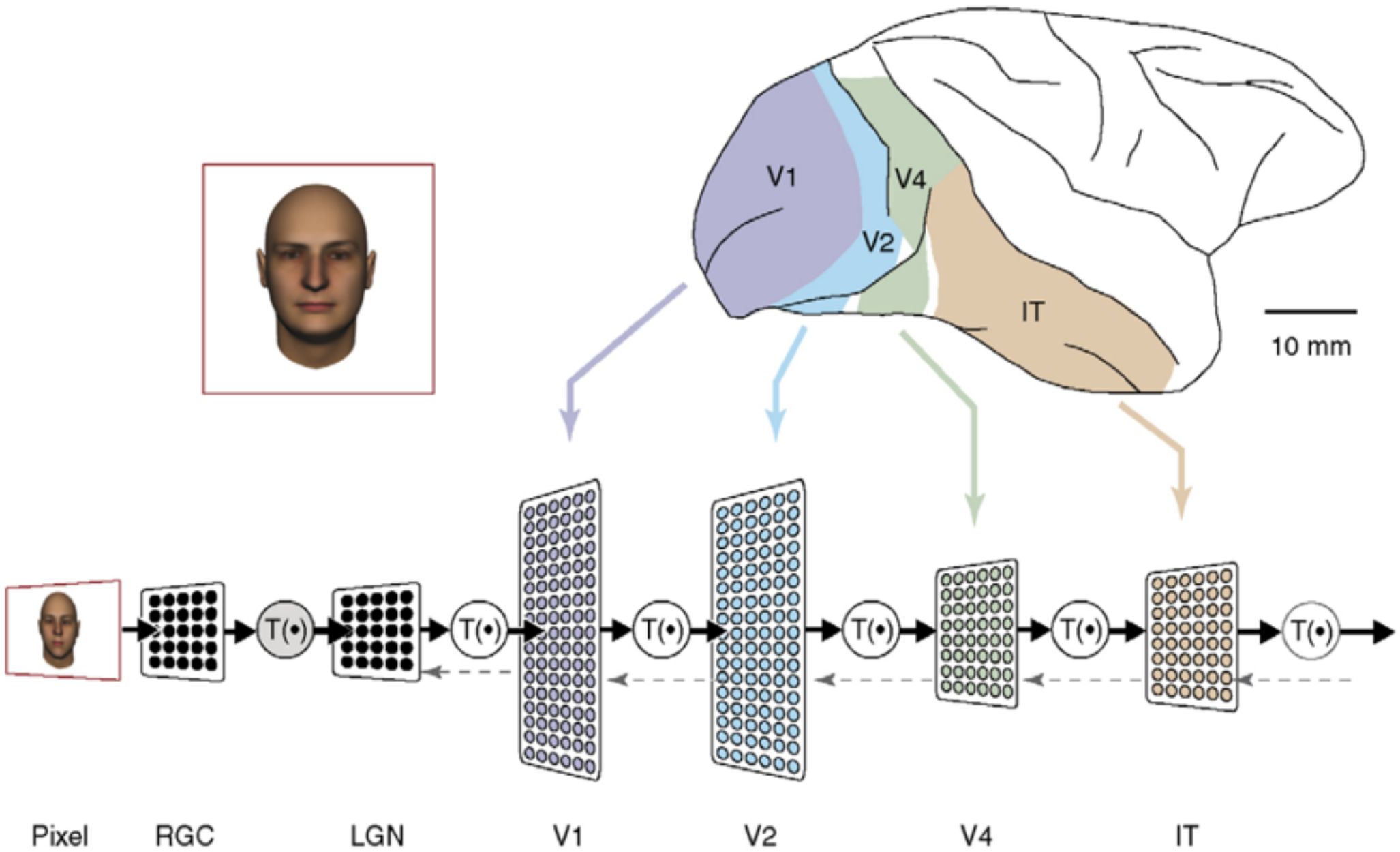
lesion “where” pathway: difficulty in spatial reasoning

lesion “what” pathway: difficulty in object recognition



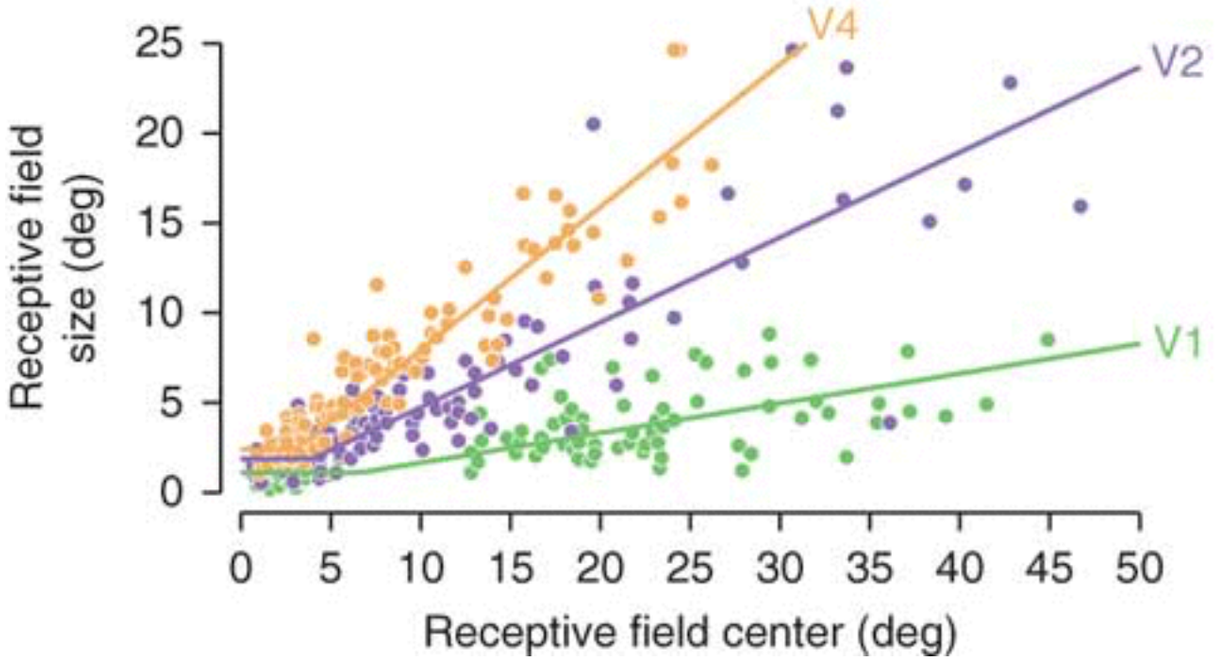
Mishkin & Ungerleider 1982

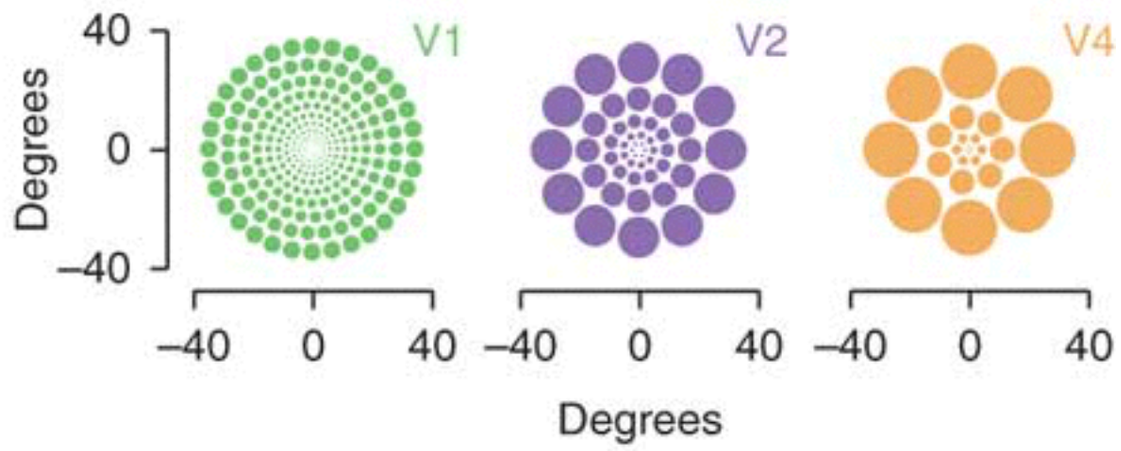
Object Recognition: The “What” Pathway



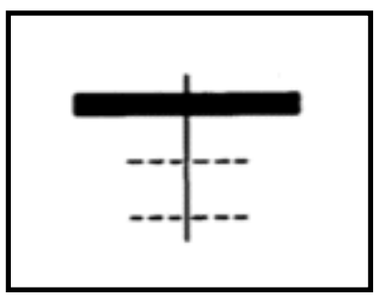
DiCarlo & Cox (2007)

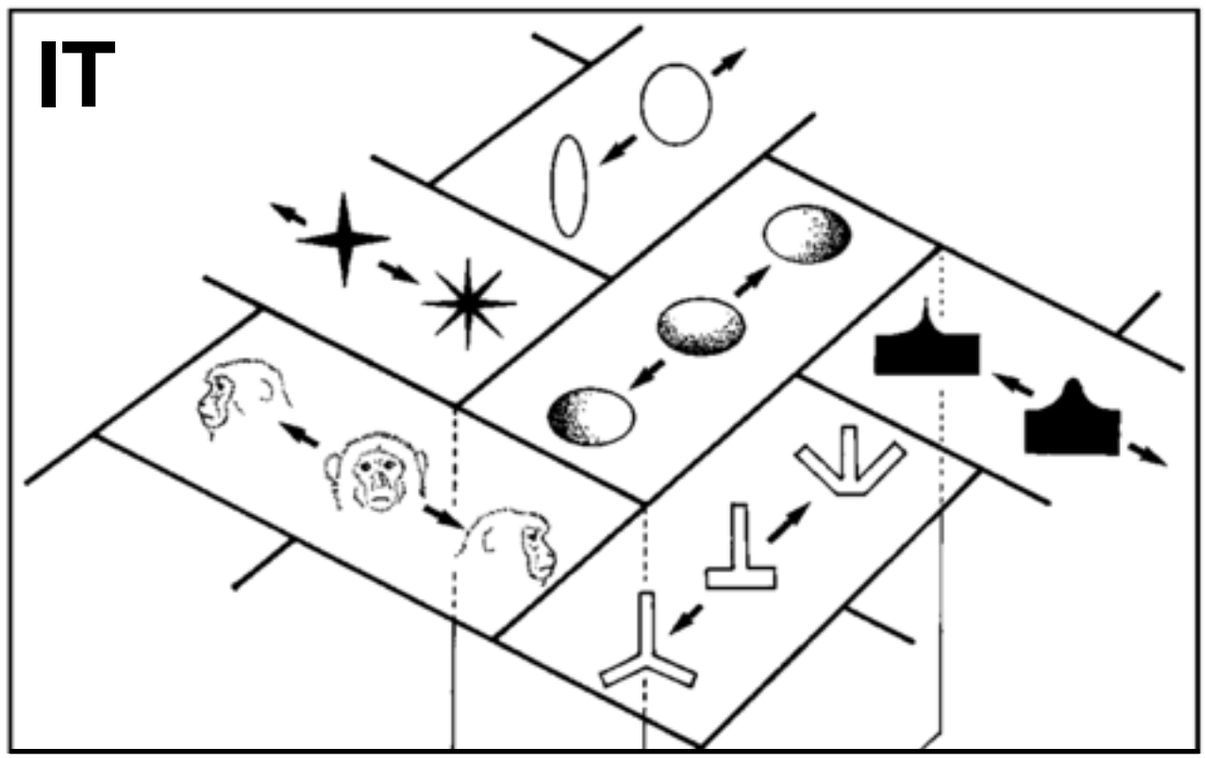
Object Recognition: The “What” Pathway





V1





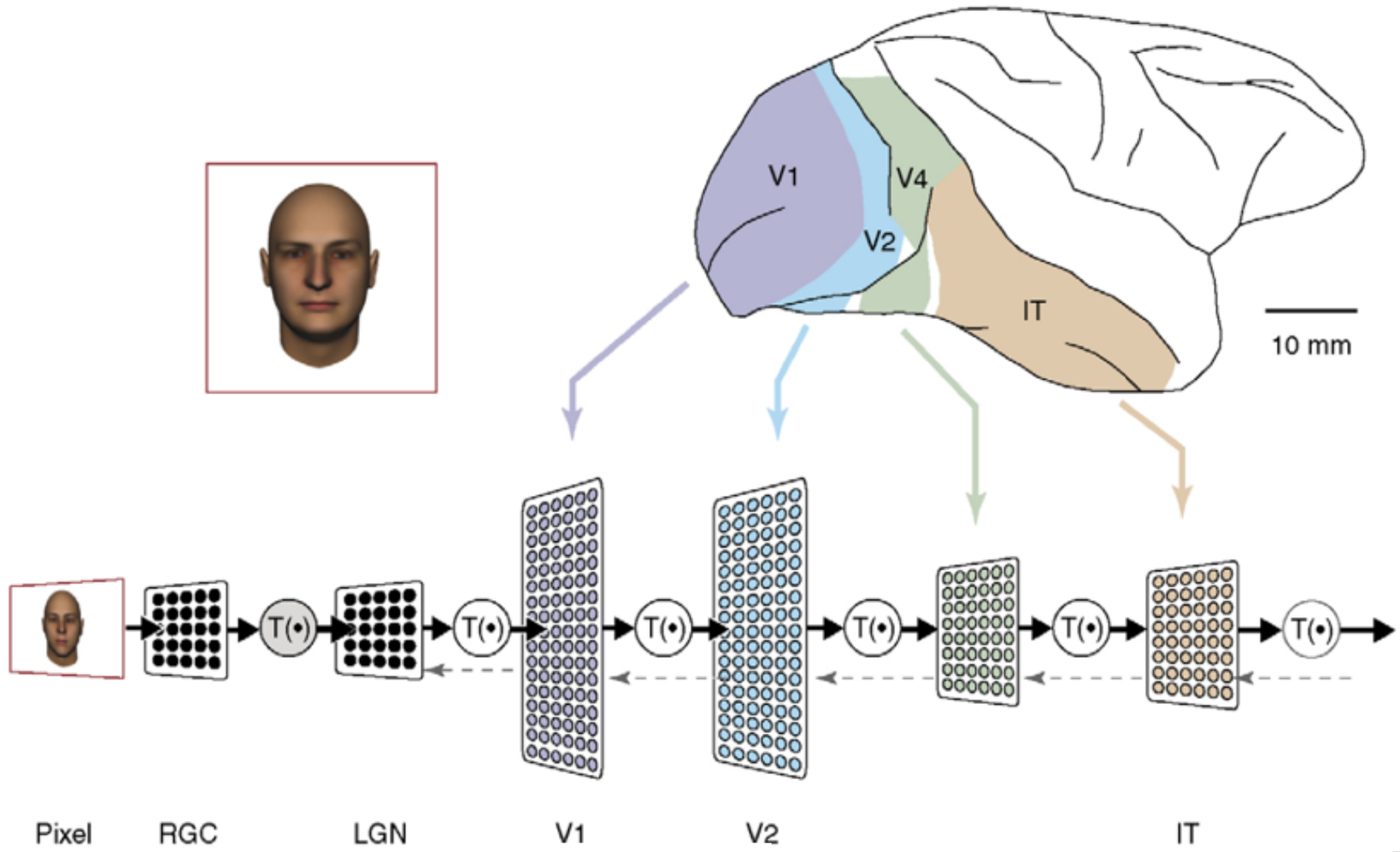
Freeman & Simoncelli (2011), Tanaka (1997)

Object Recognition: Building Features and Invariance

visual processing is done in stages

each area performs a transformation on its inputs

invariance is built gradually across many successive steps

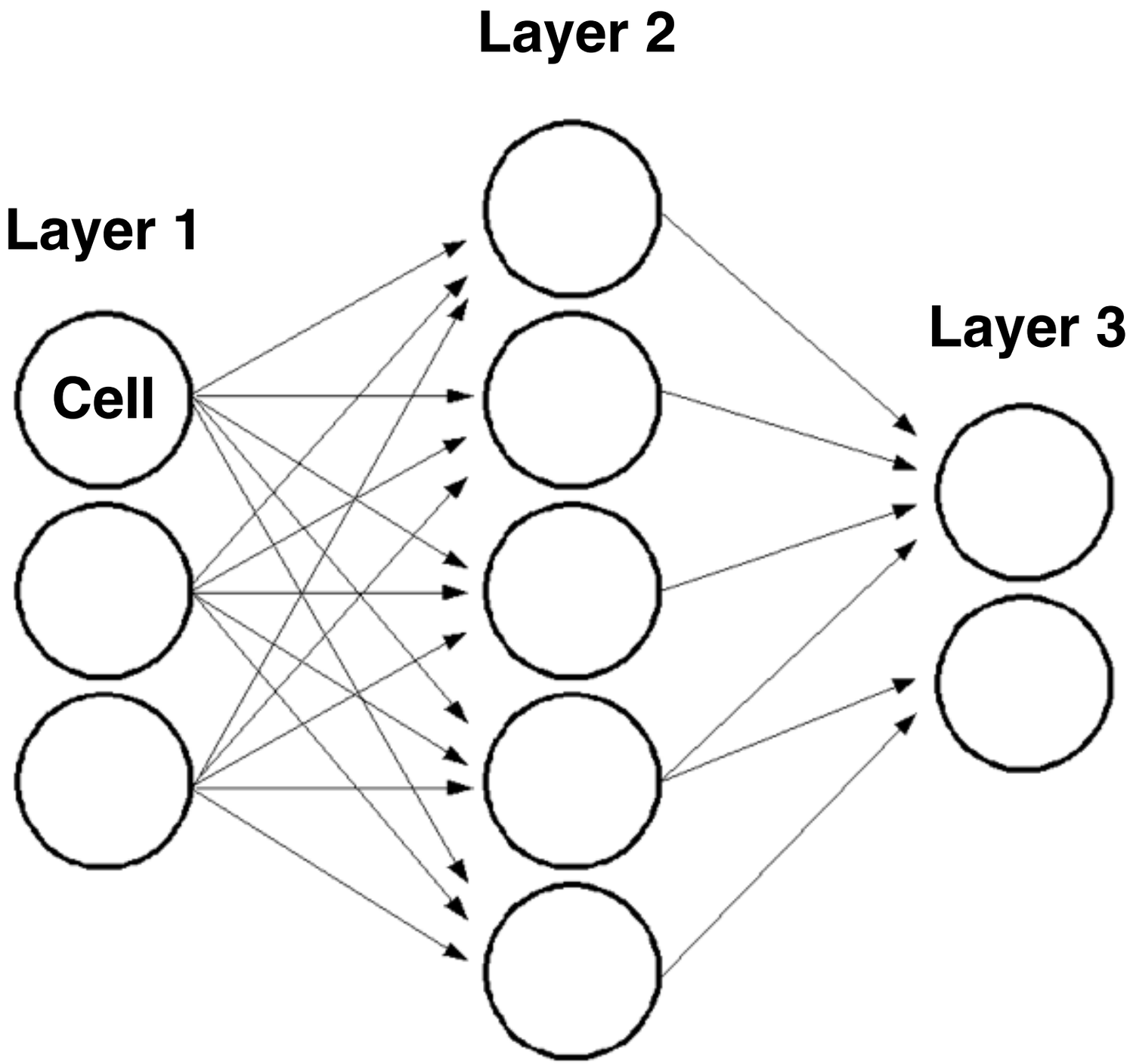


DiCarlo & Cox (2007)

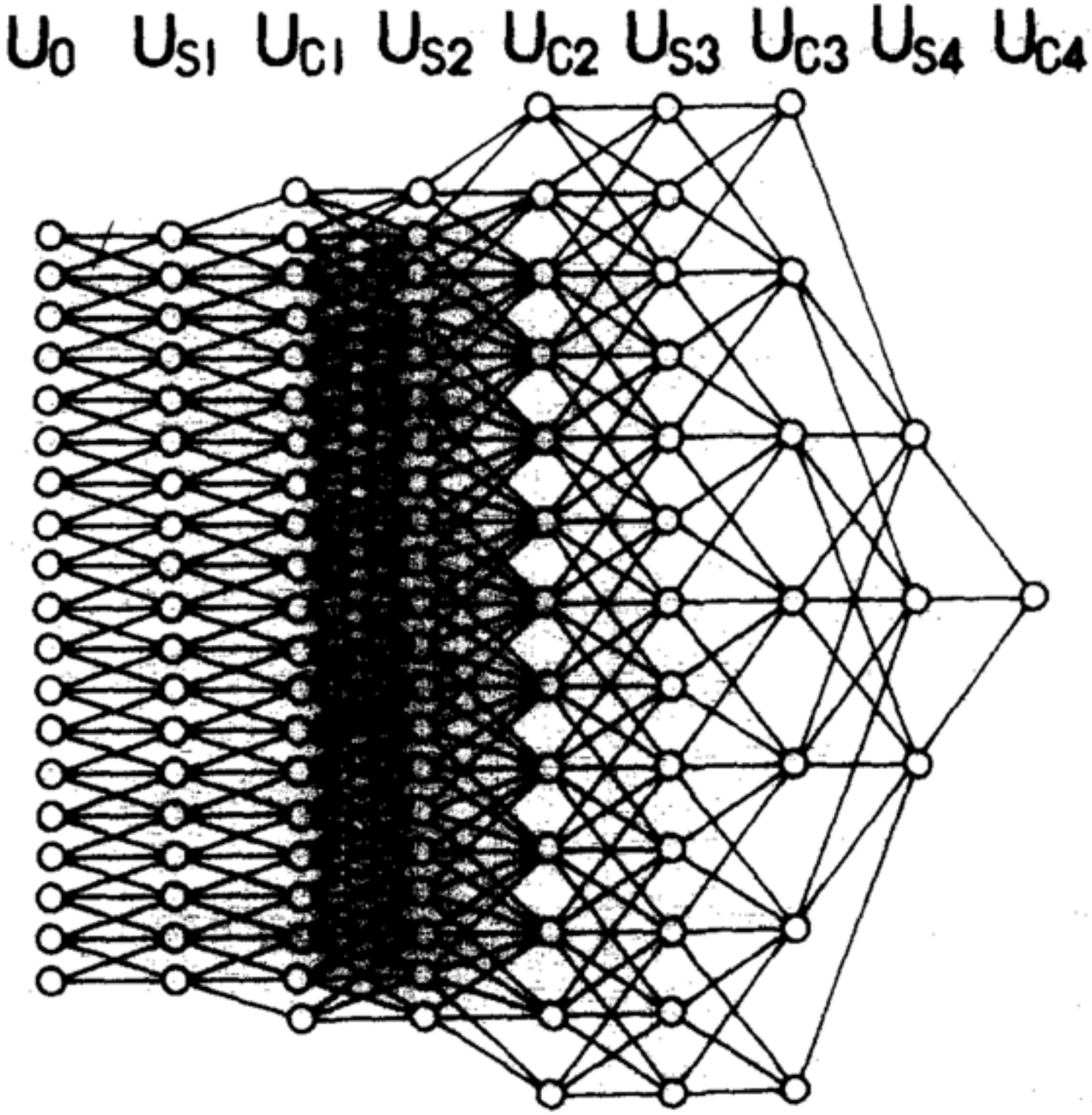
2. Hierarchical Pattern Recognition Systems

neuroscience-inspired computer vision

Neocognitron: Neural Network



Neocognitron: Neural Network



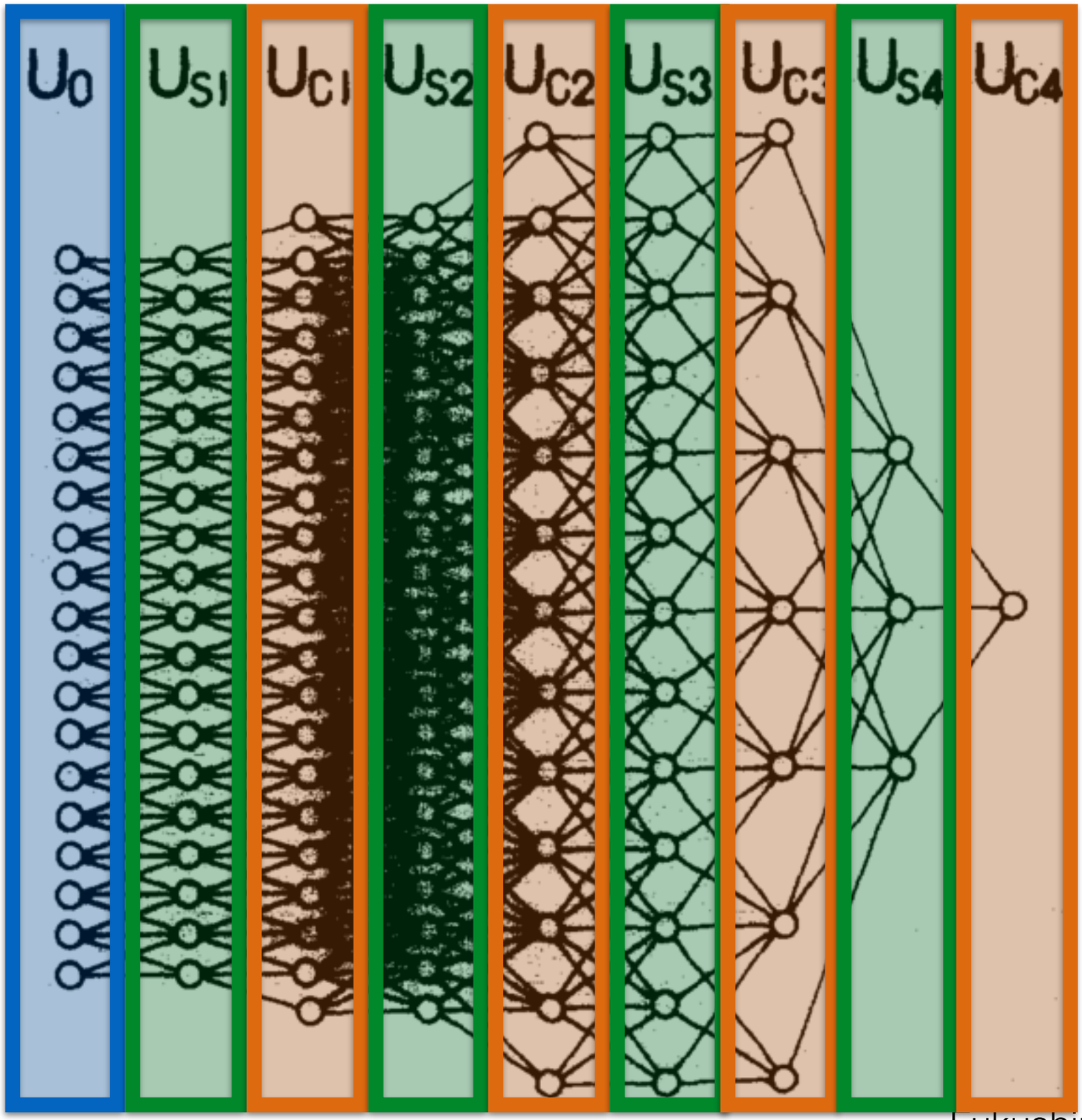
Fukushima (1988)

Neocognitron: Neural Network

input layer

S-cell layer

C-cell layer



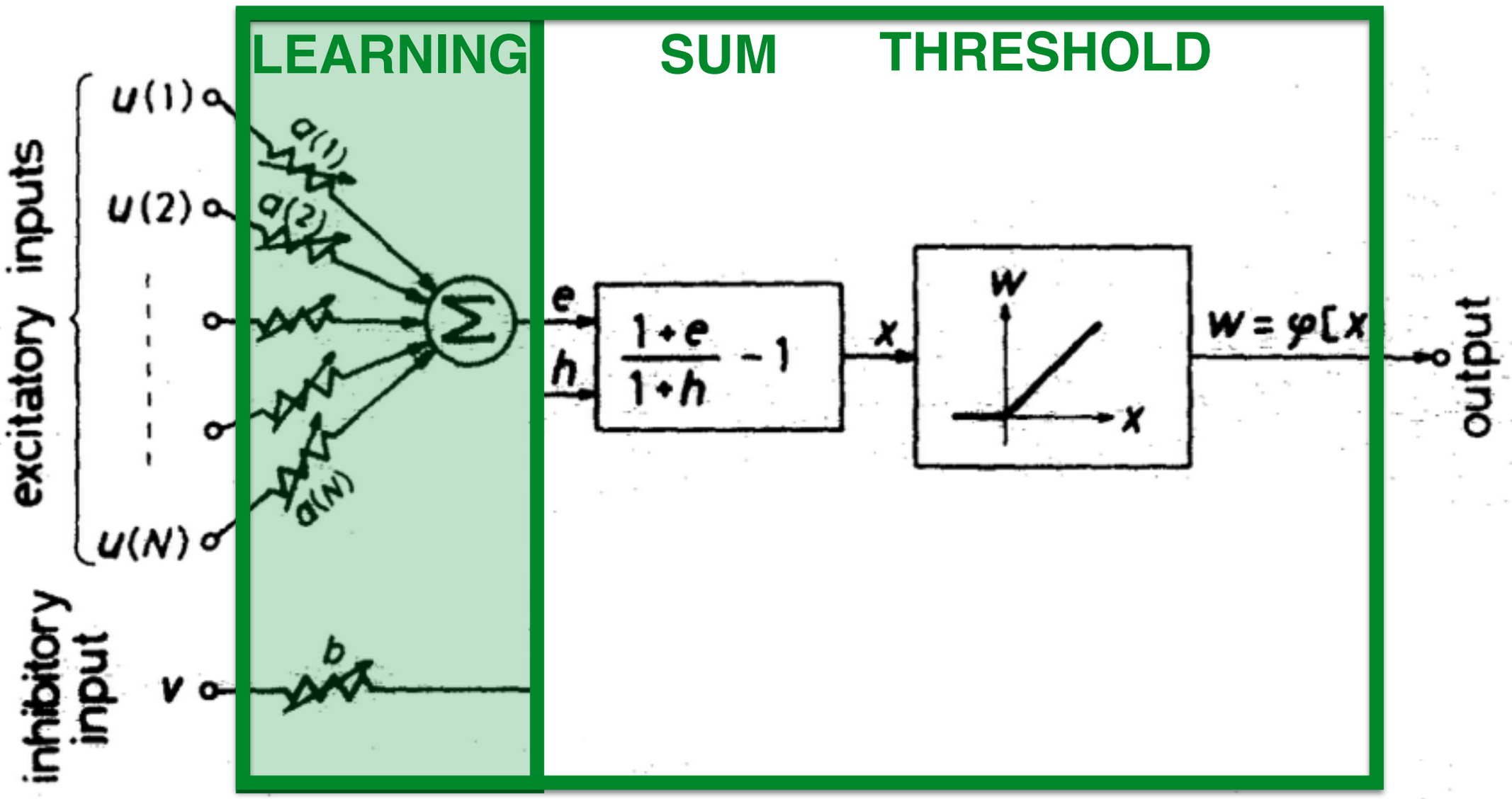
Fukushima (1988)

Neocognitron: S-Cell

C-cell layer / input layer

S-cell layer

C-cell layer



Fukushima (1988)

Neocognitron: S-Cell

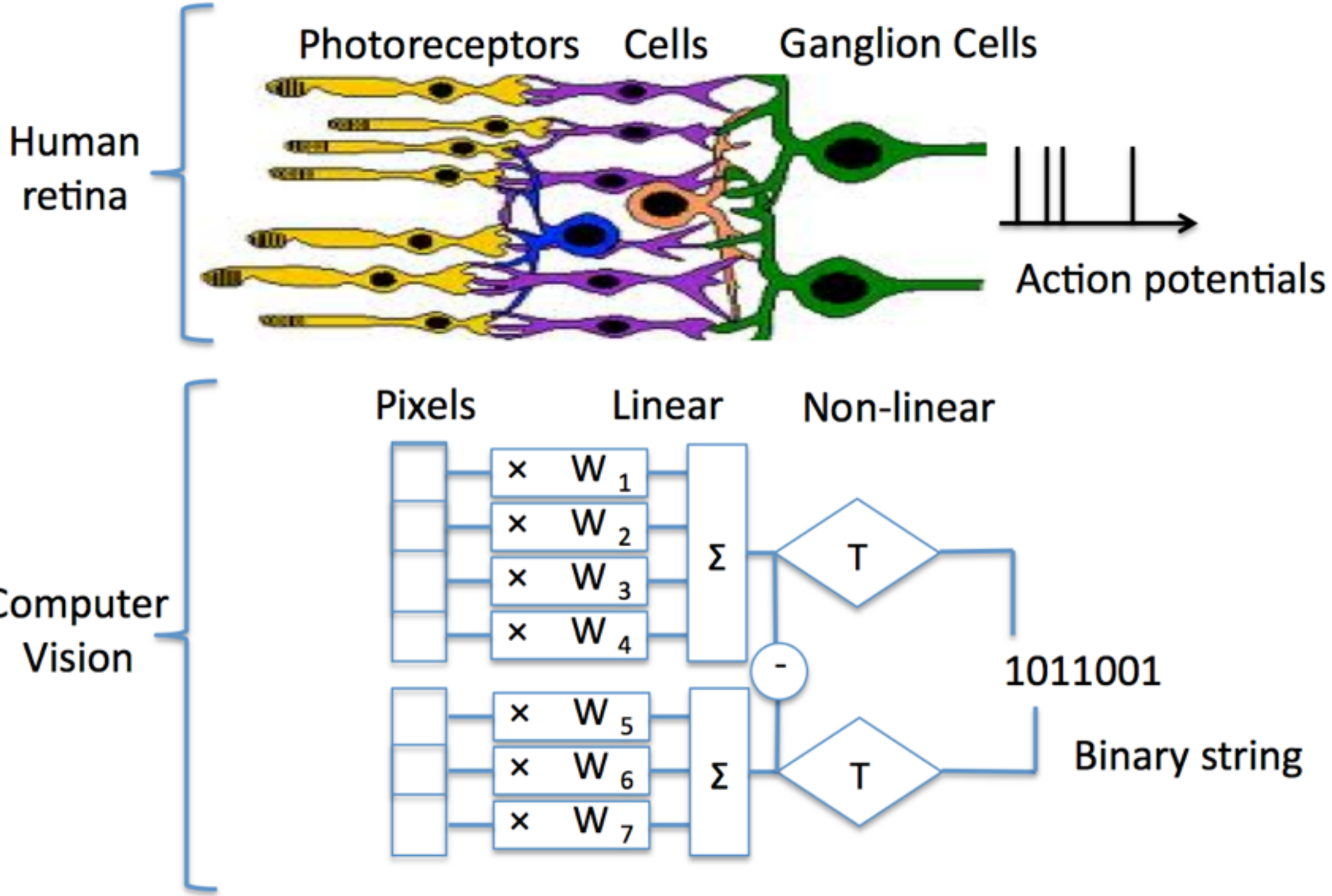
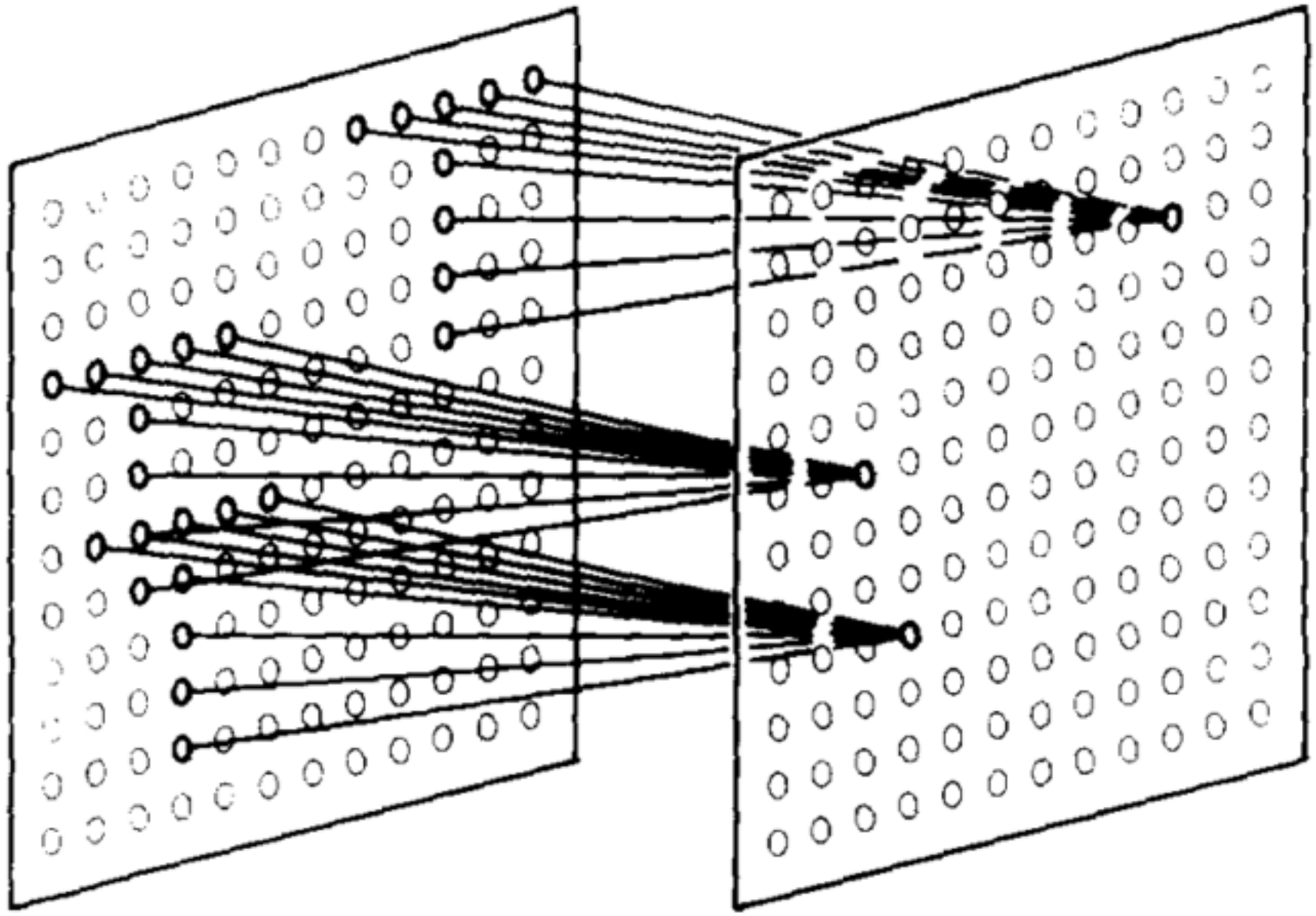


figure courtesy of A. Alahi

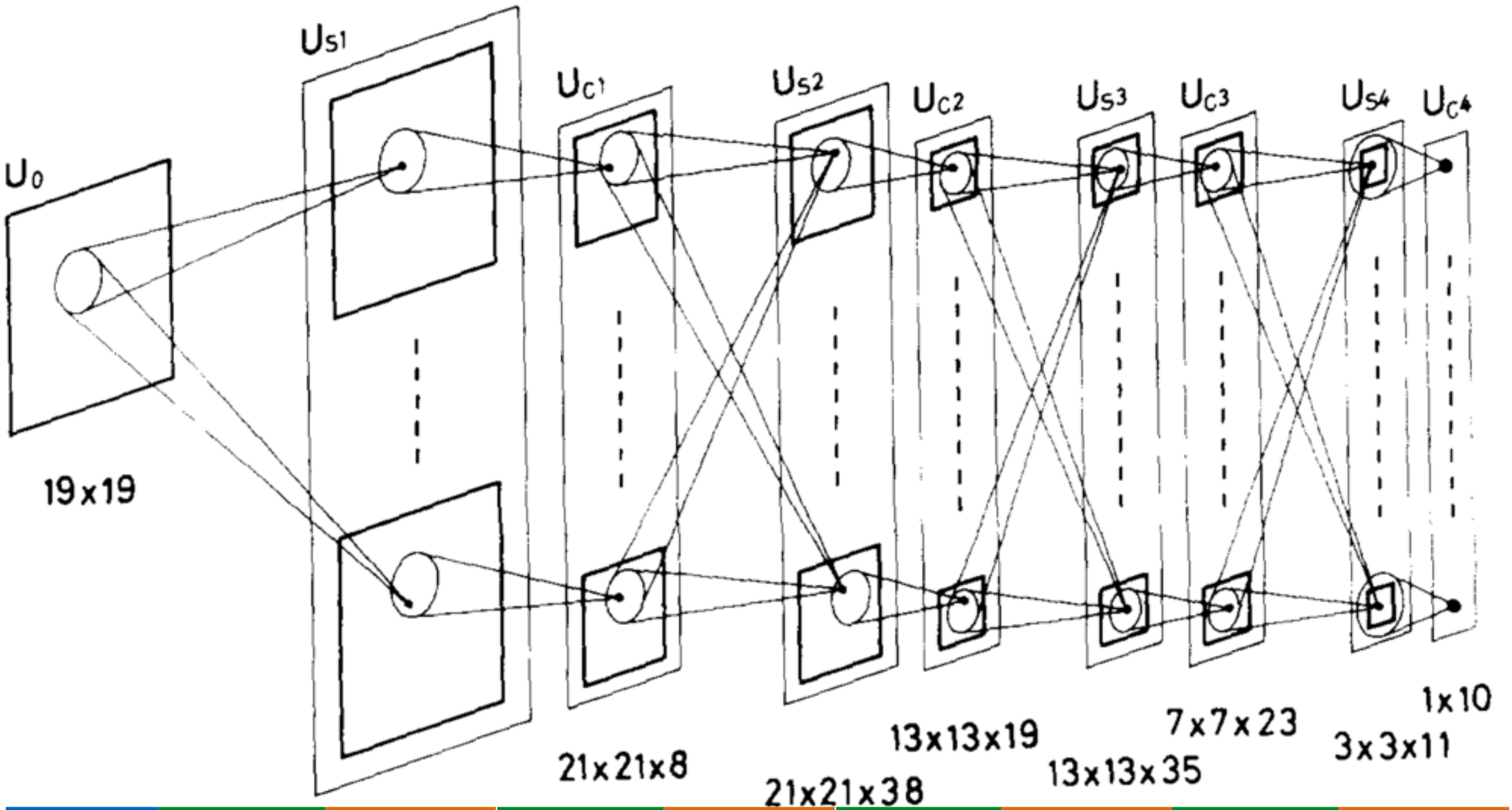
Neocognitron: C-Cell "Pooling"

building position invariance



Fukushima (1988)

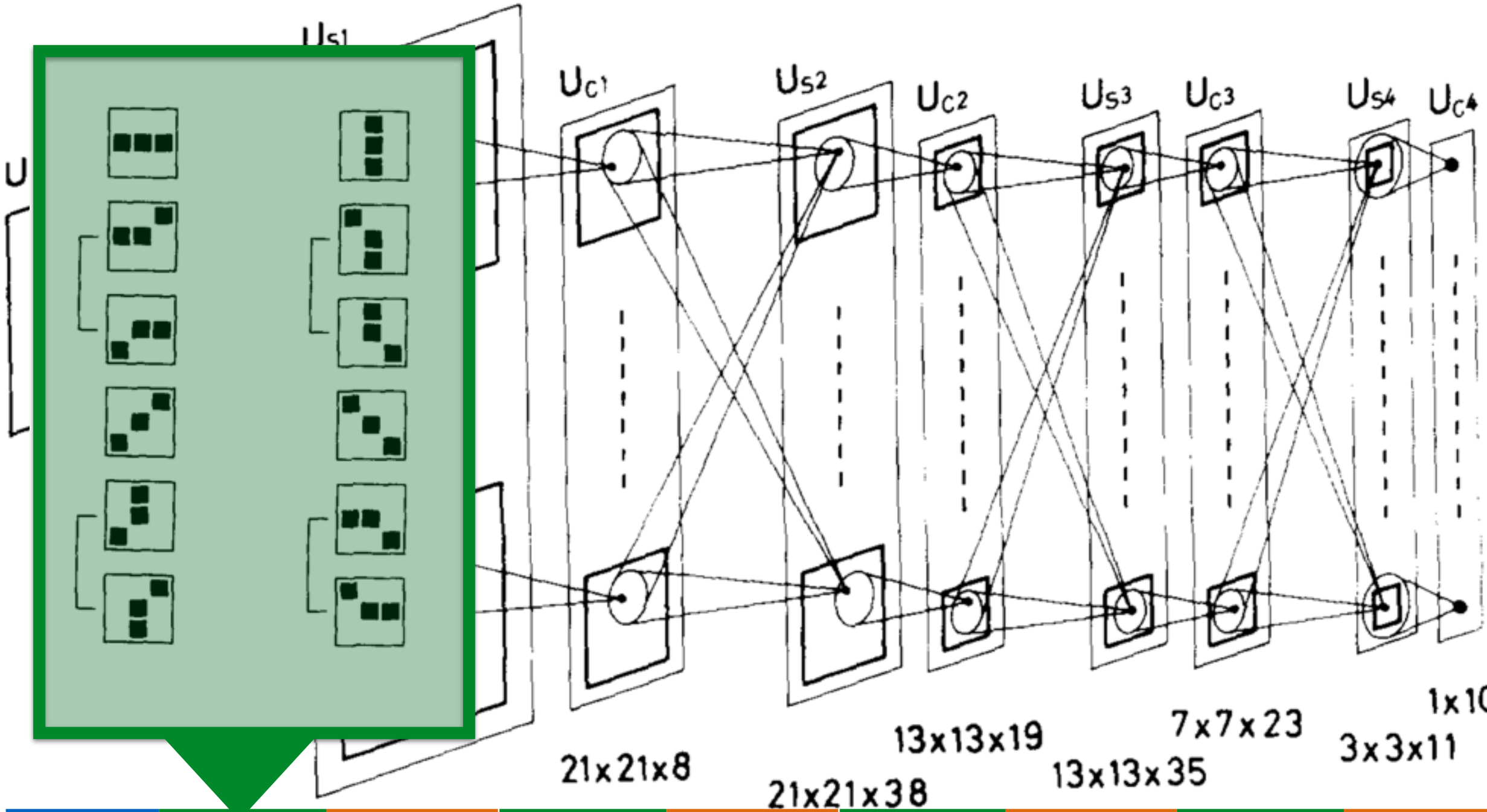
Neocognitron: Network



input	S-cell	C-cell	S-cell	C-cell	S-cell	C-cell	S-cell	C-cell

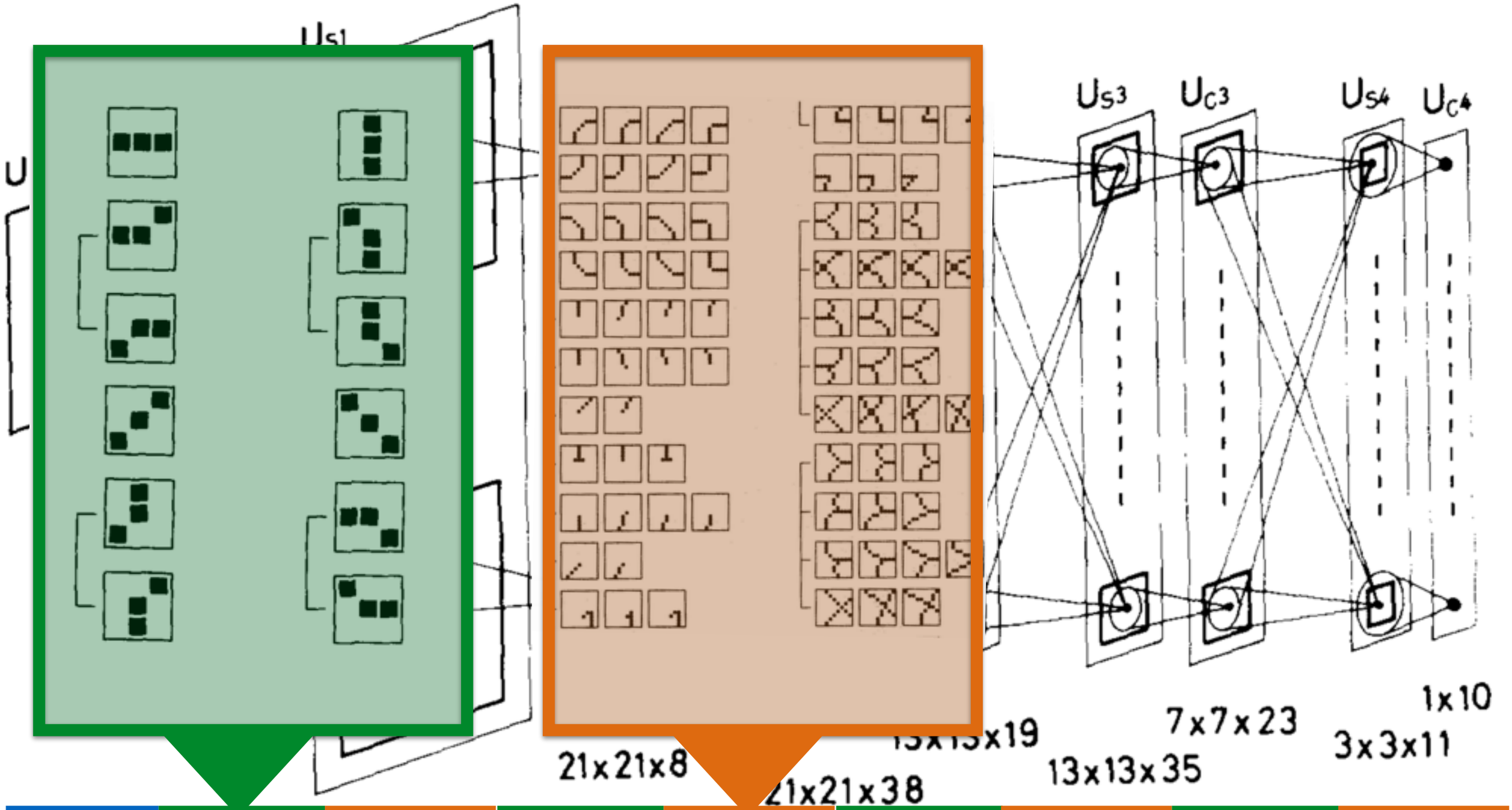
Fukushima (1988)

Neocognitron: Network



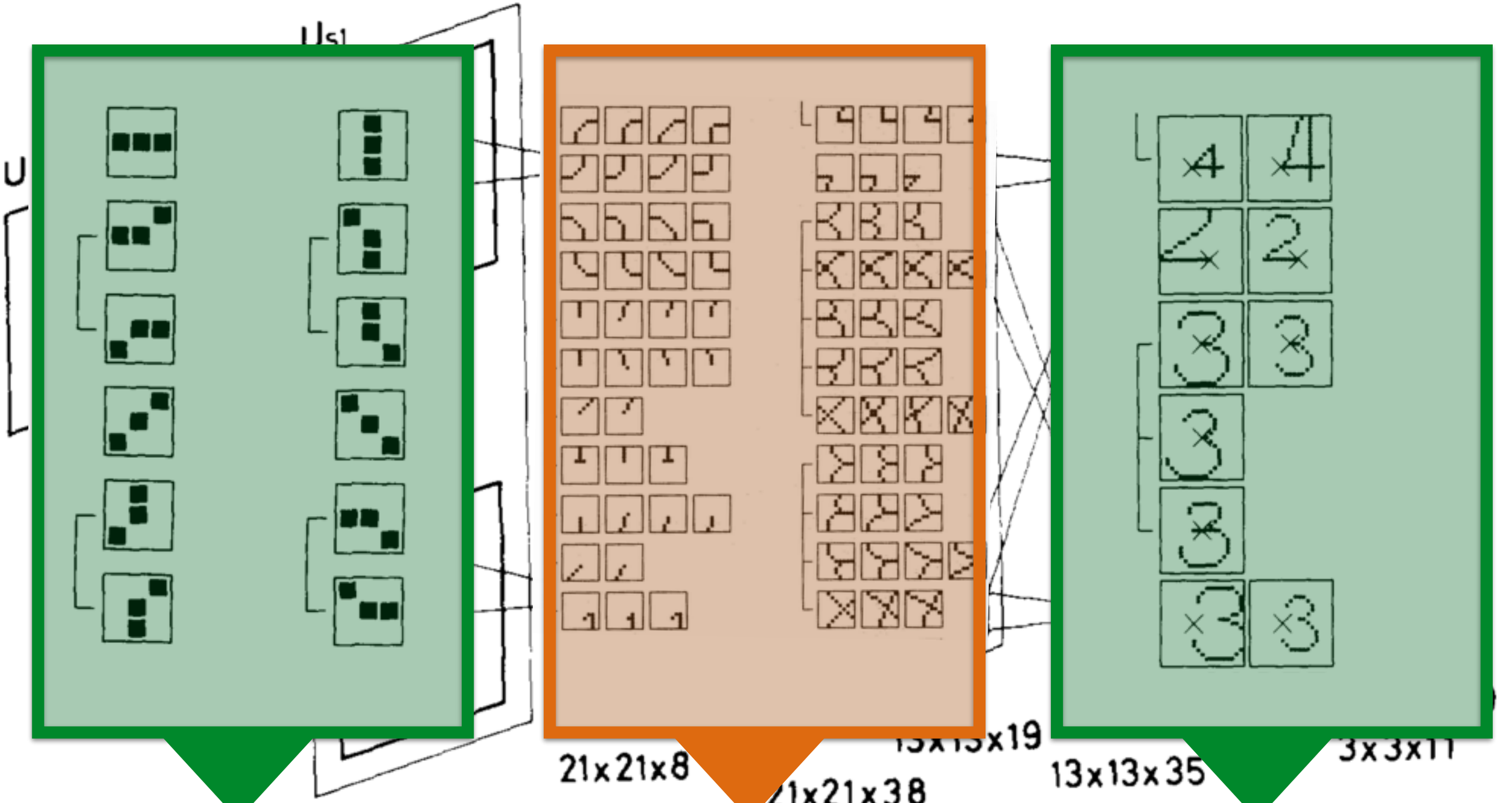
Fukushima (1988)

Neocognitron: Network



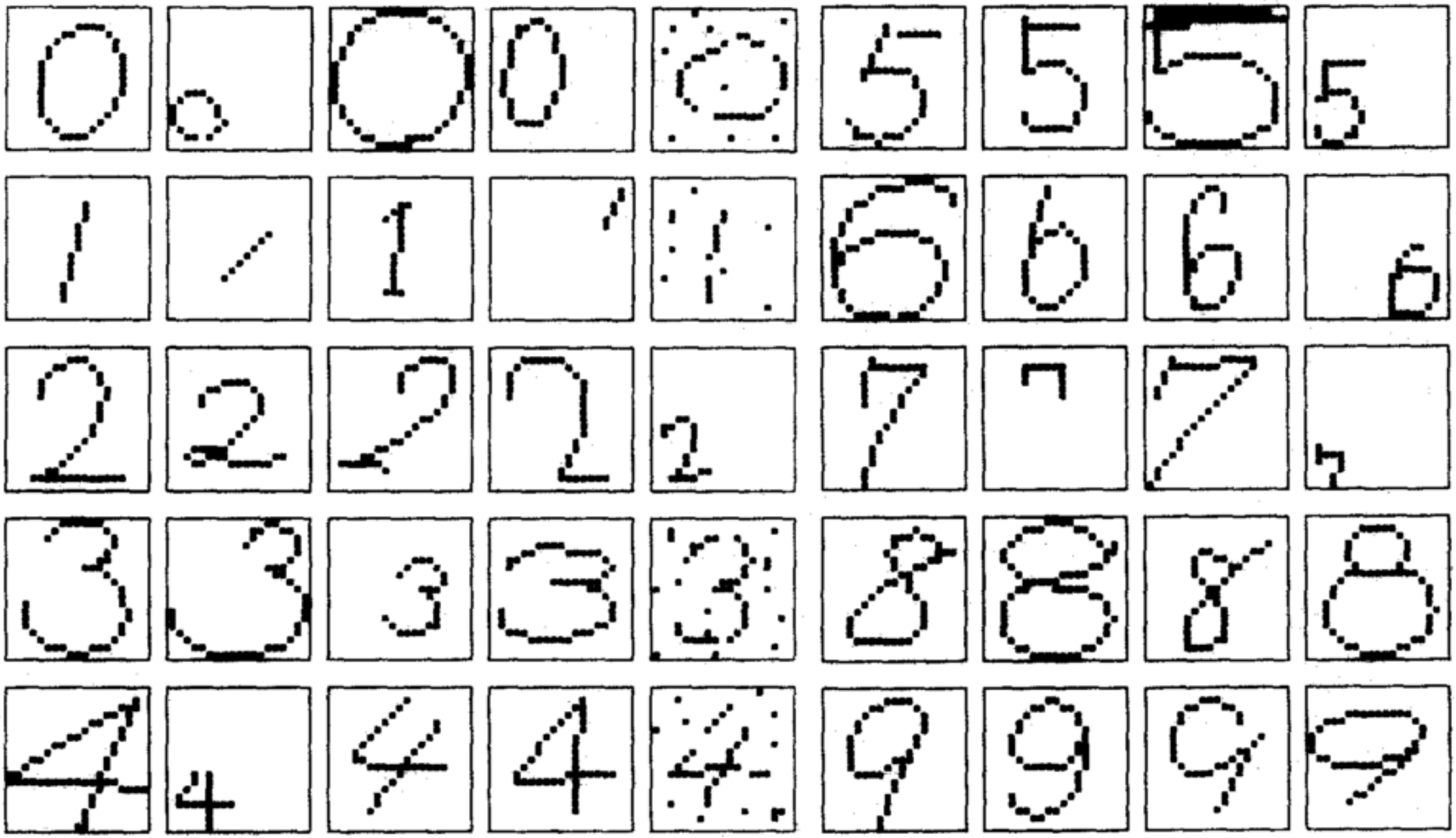
Fukushima (1988)

Neocognitron: Network



Fukushima (1988)

Neocognitron: Robust Results



Fukushima (1988)

Neocognitron

biologically inspired hierarchical processing pipeline

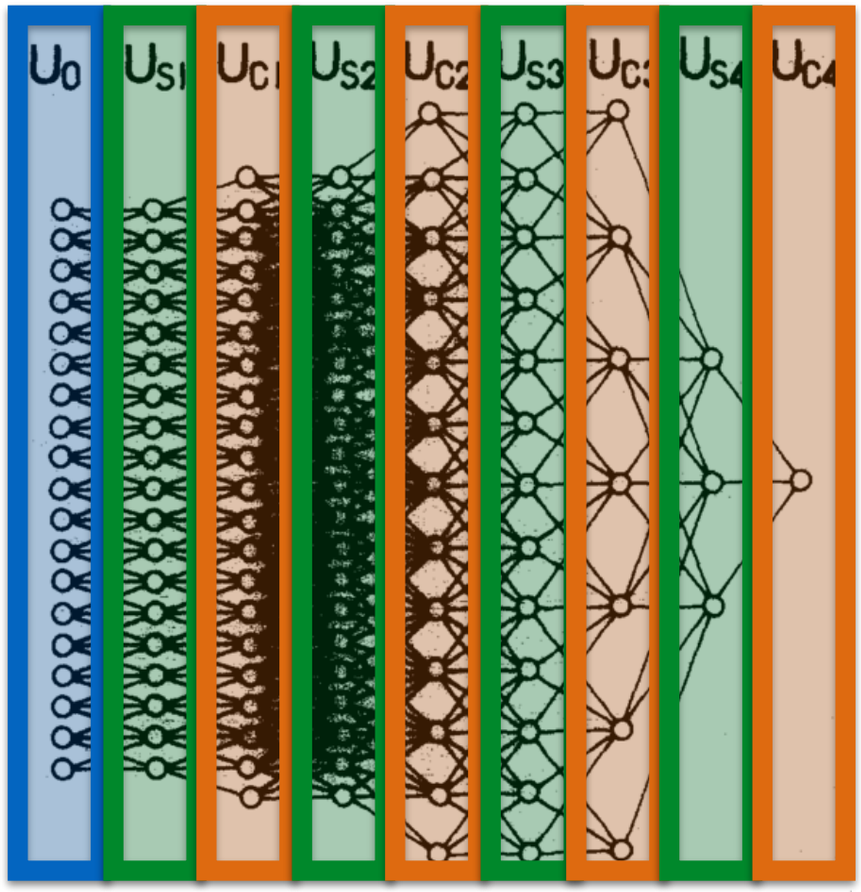
invariance is built gradually across many successive steps

simple neural network solves complicated non-linear problem

input layer

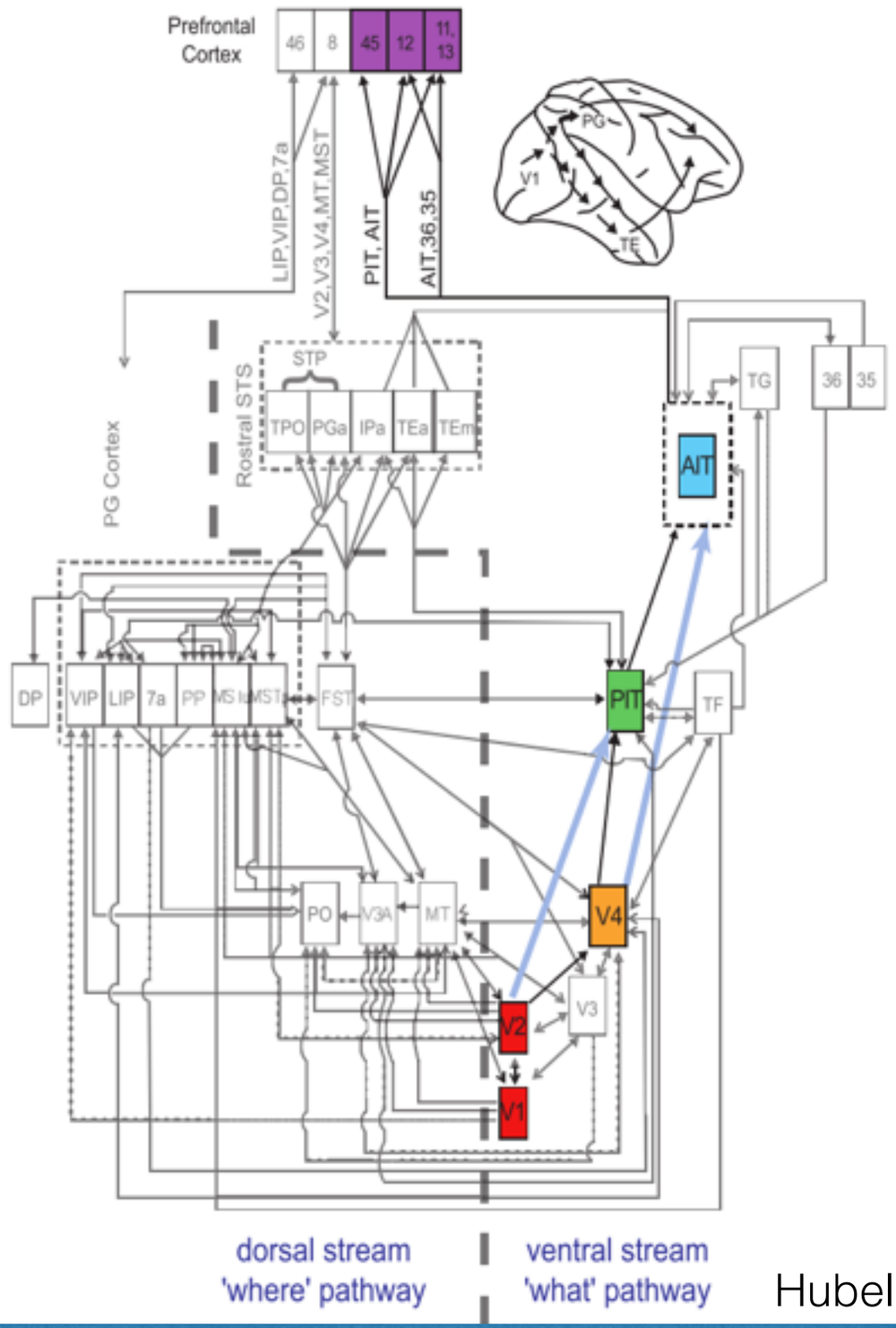
S-cell layer

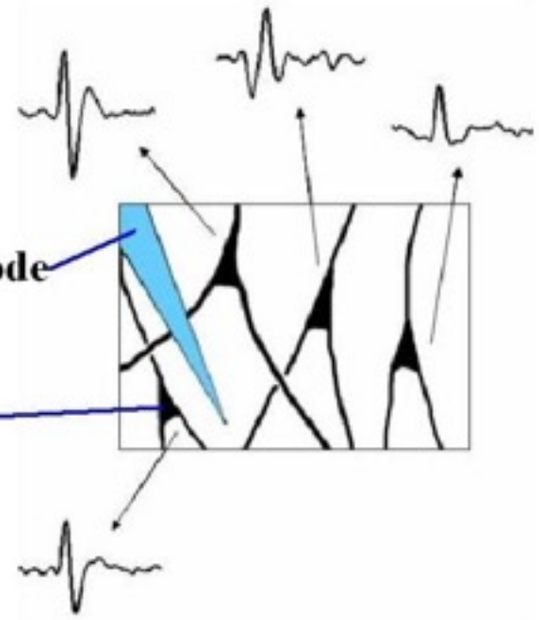
C-cell layer



Fukushima (1988)

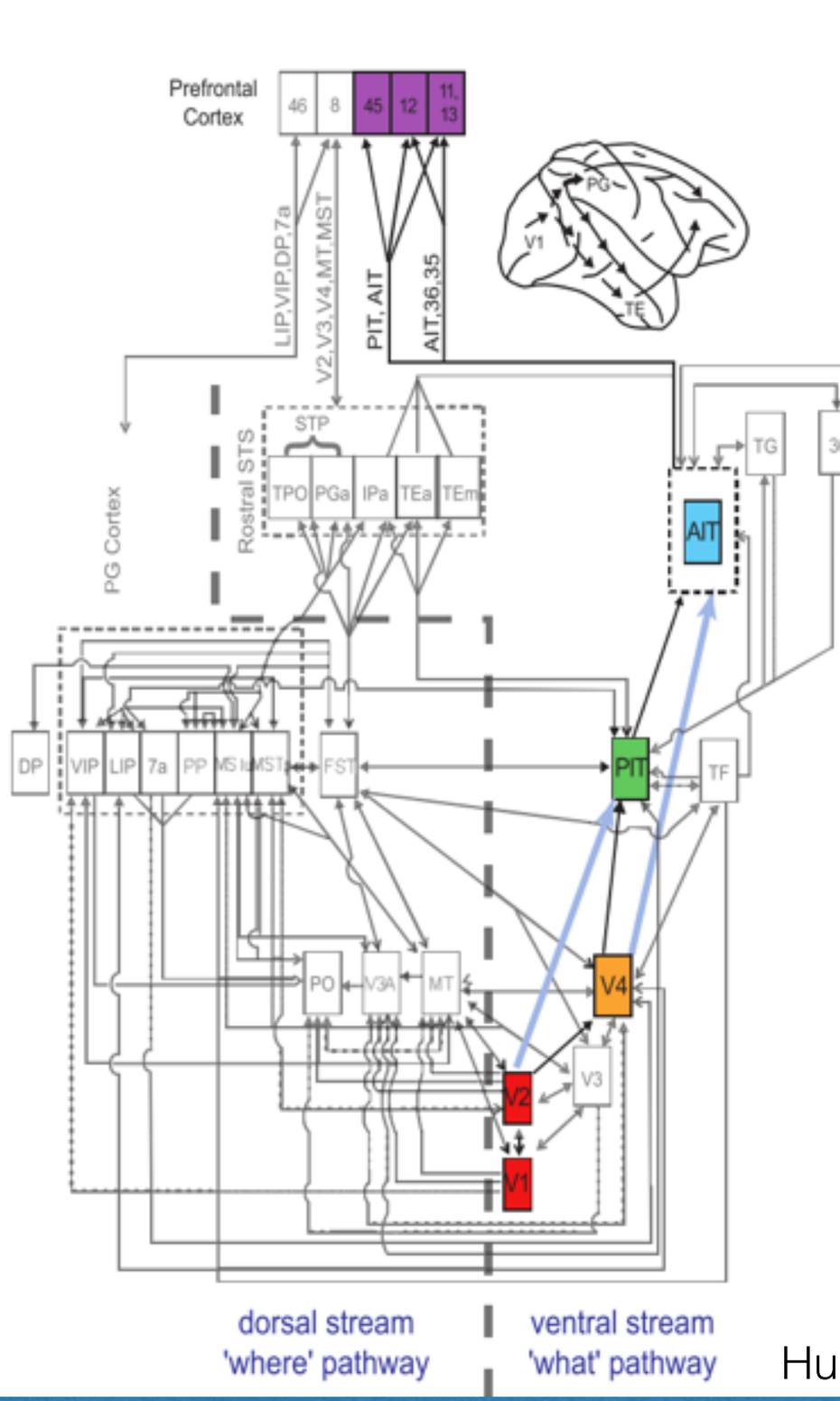
Feed-Forward Object Recognition Model

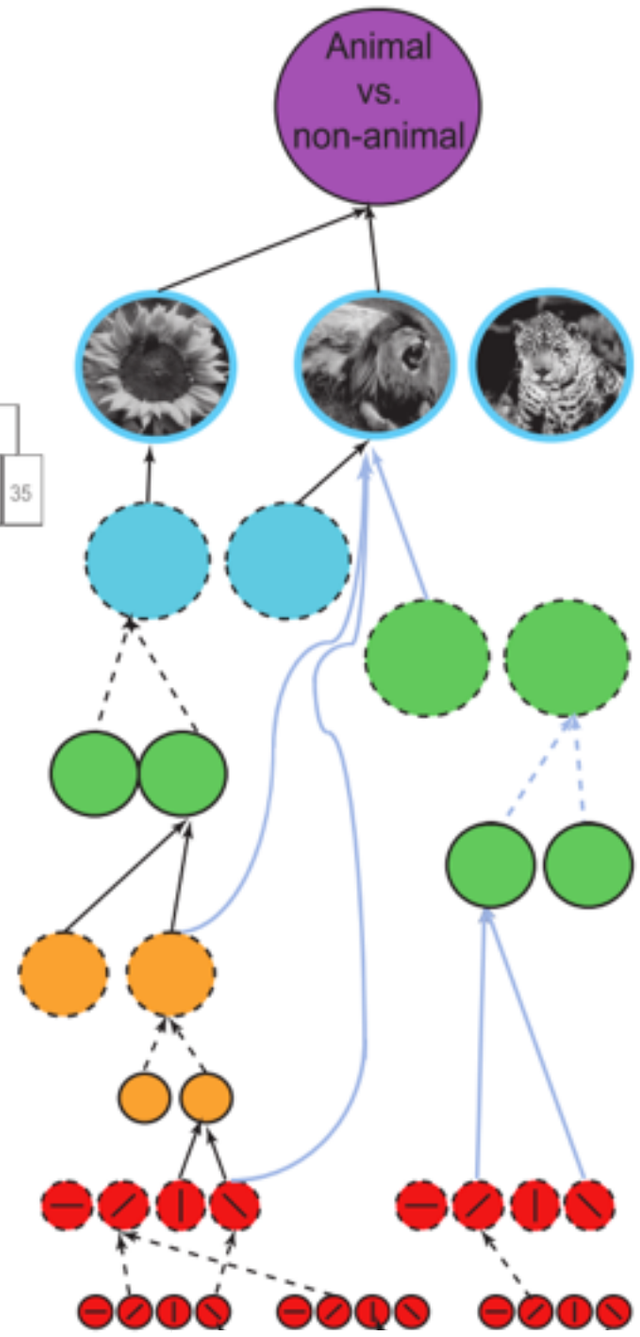


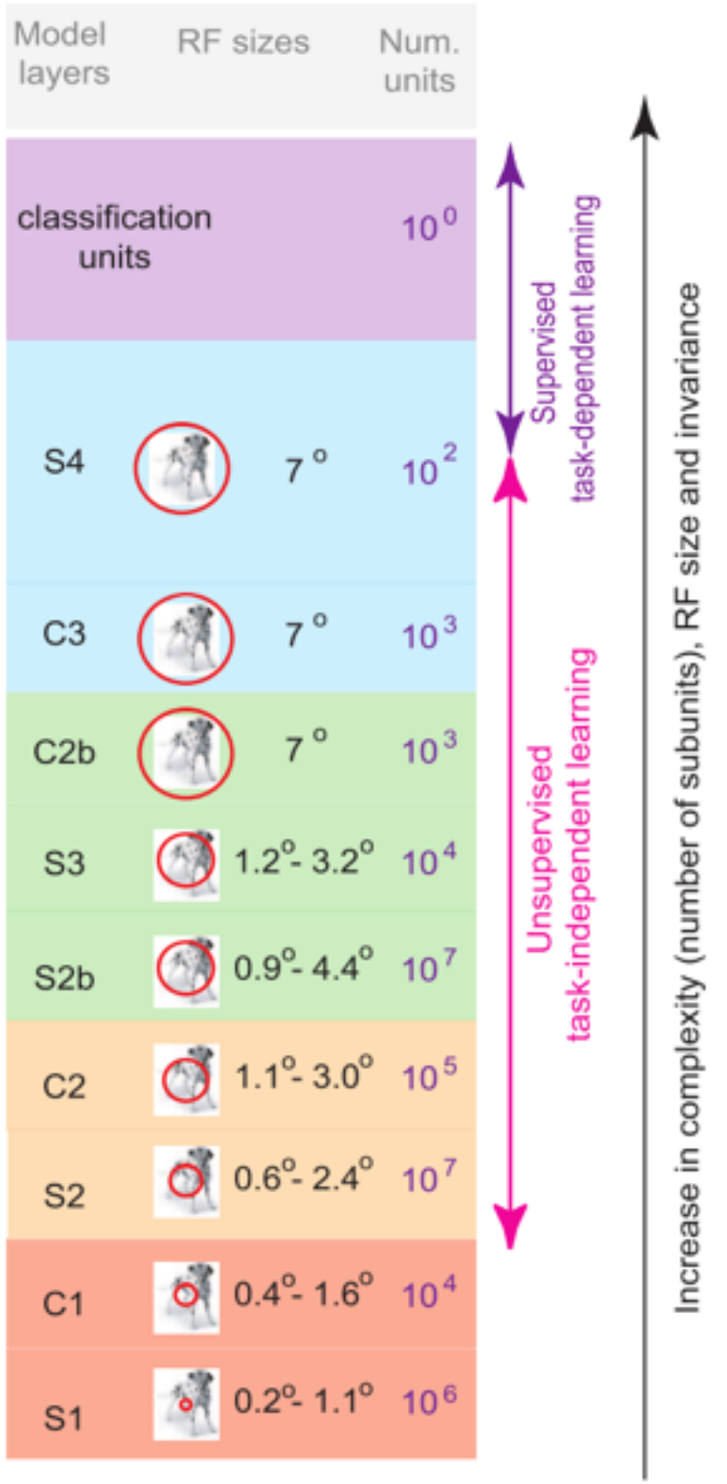


Hubel & Wiesel (1962), Riesenhuber & Poggio (1999), Serre et al. (2007)

Feed-Forward Object Recognition Model



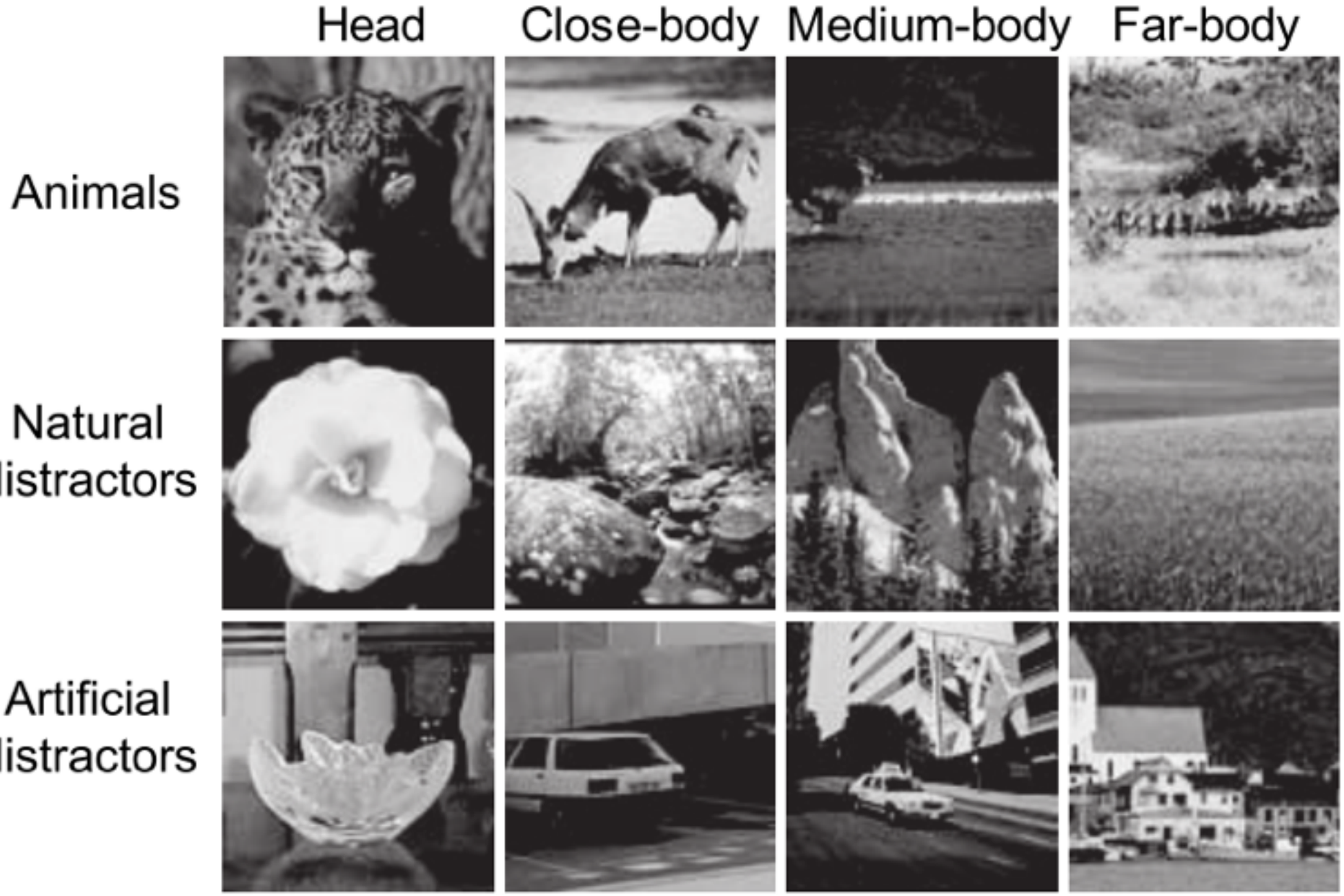




Hubel & Wiesel (1962), Riesenhuber & Poggio (1999), Serre et al. (2007)

Feed-Forward Object Recognition Model

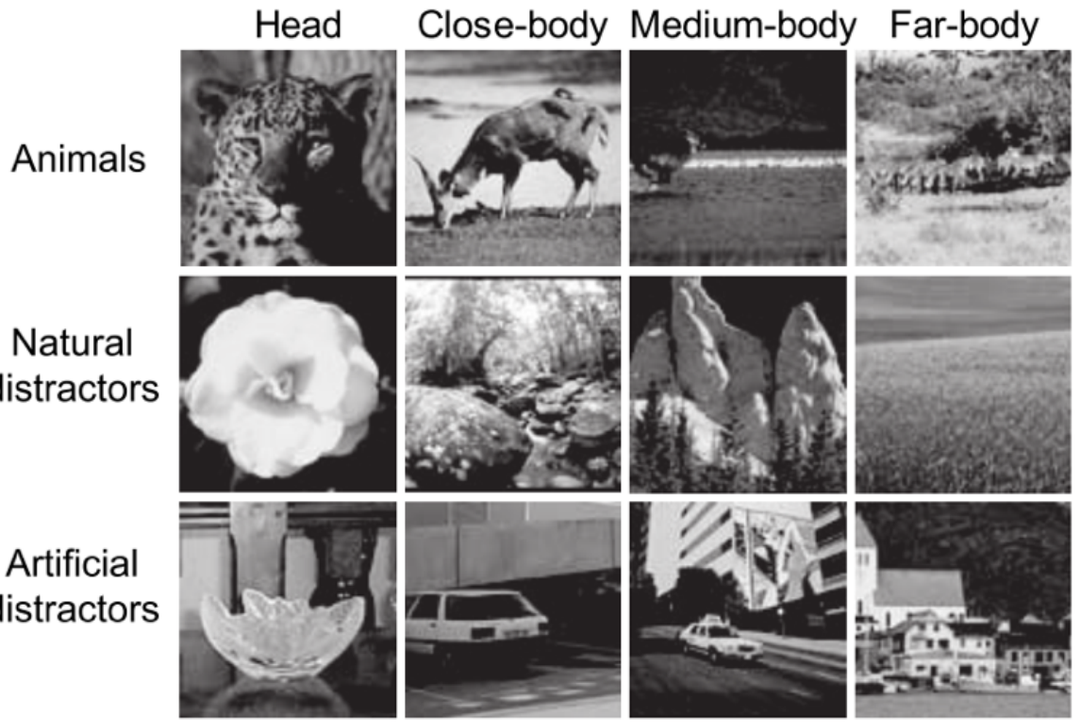
task: is there an animal in the picture?

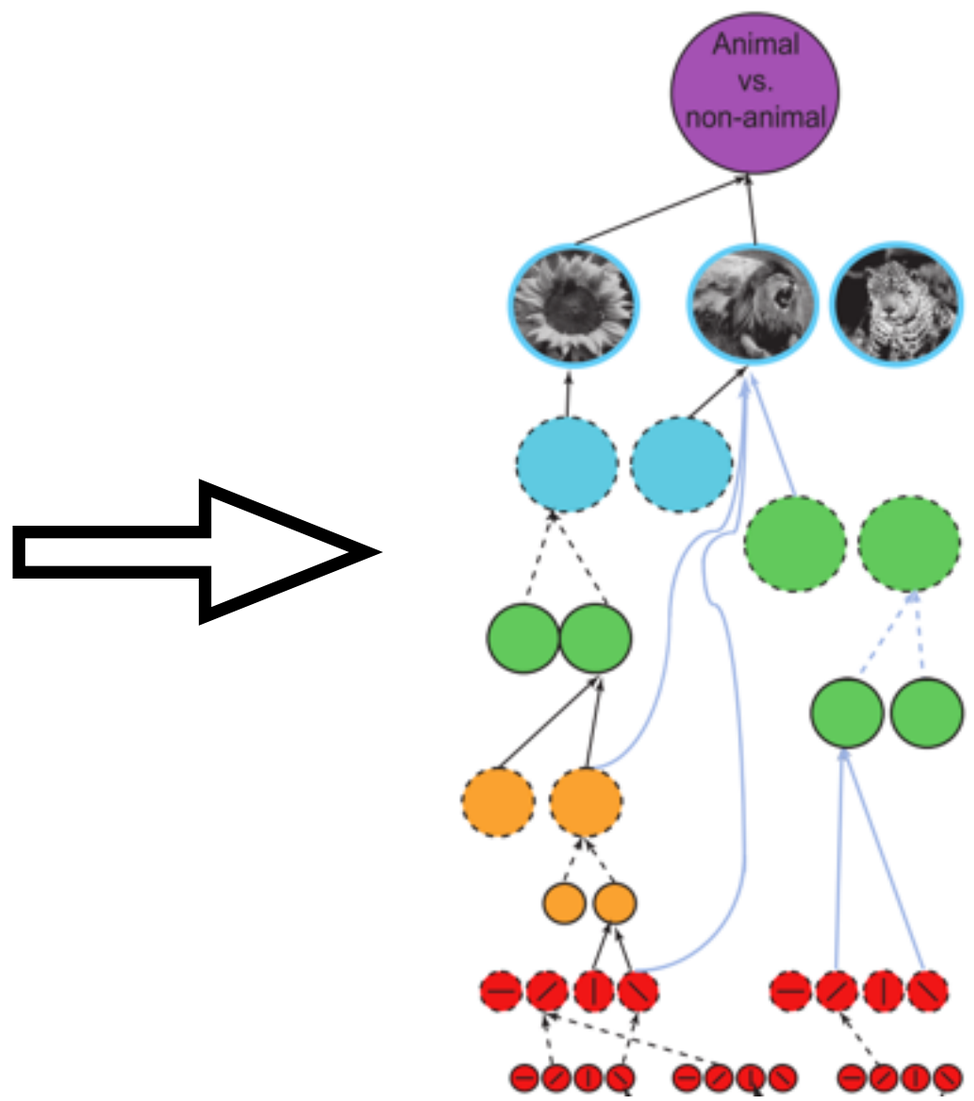


Serre et al. (2007)

Feed-Forward Object Recognition Model

task: is there an animal in the picture?

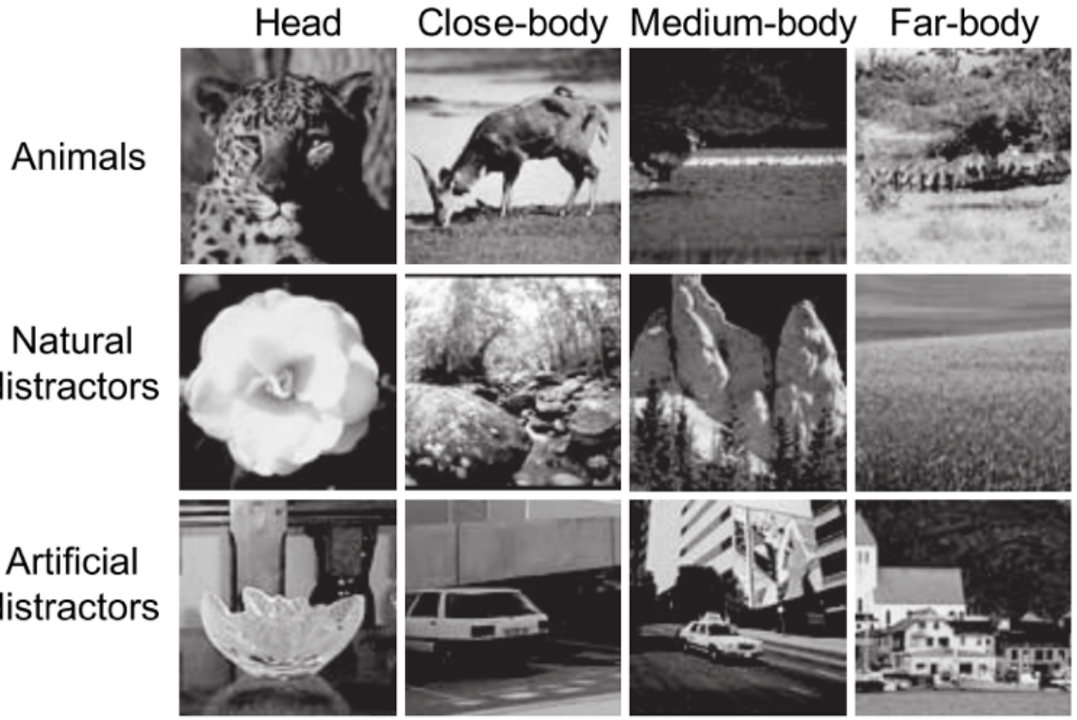


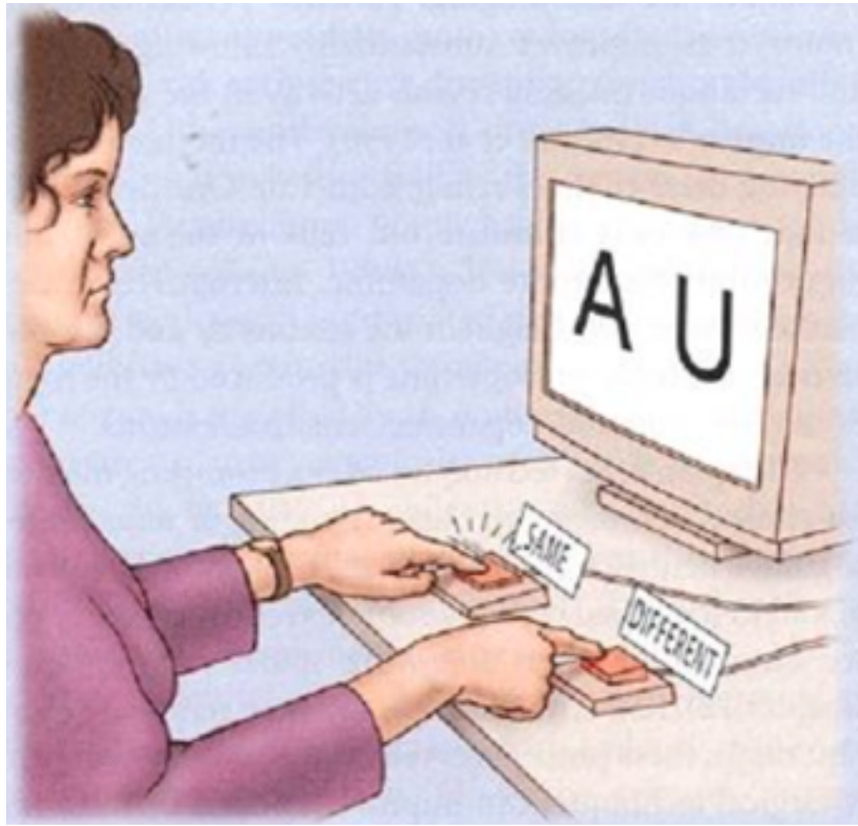


Serre et al. (2007)

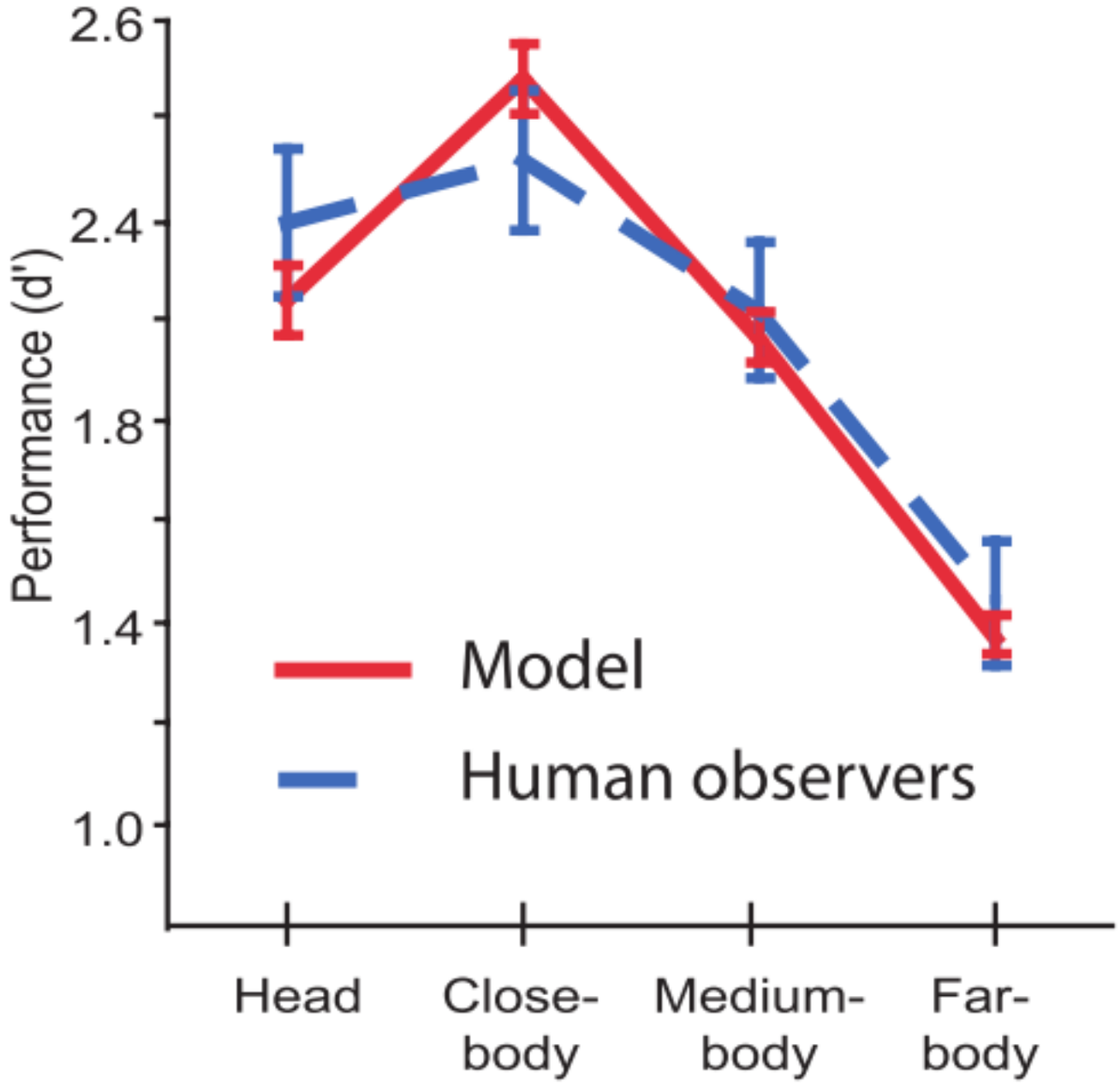
Feed-Forward Object Recognition Model

task: is there an animal in the picture?



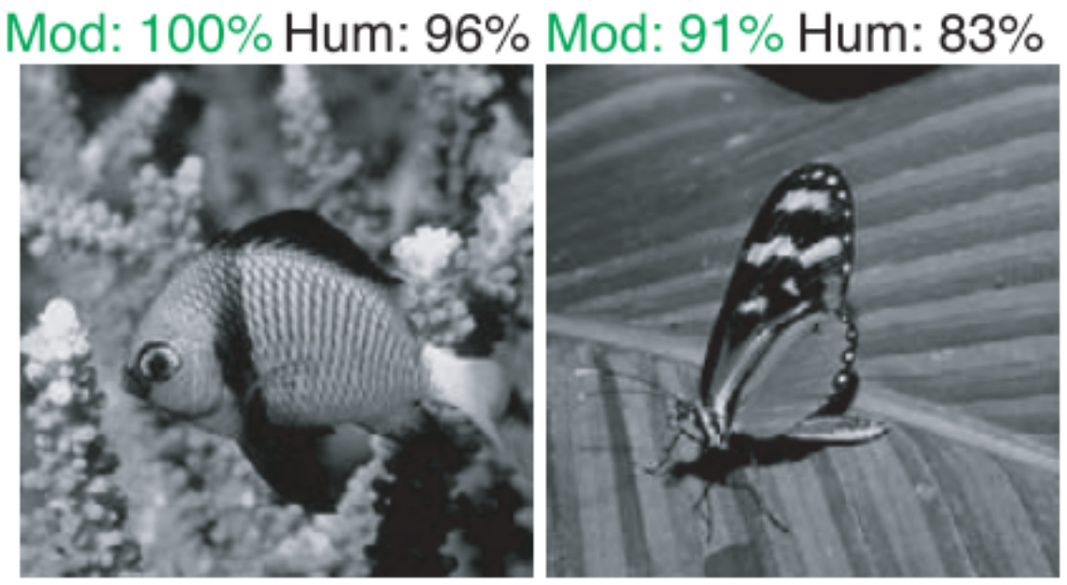


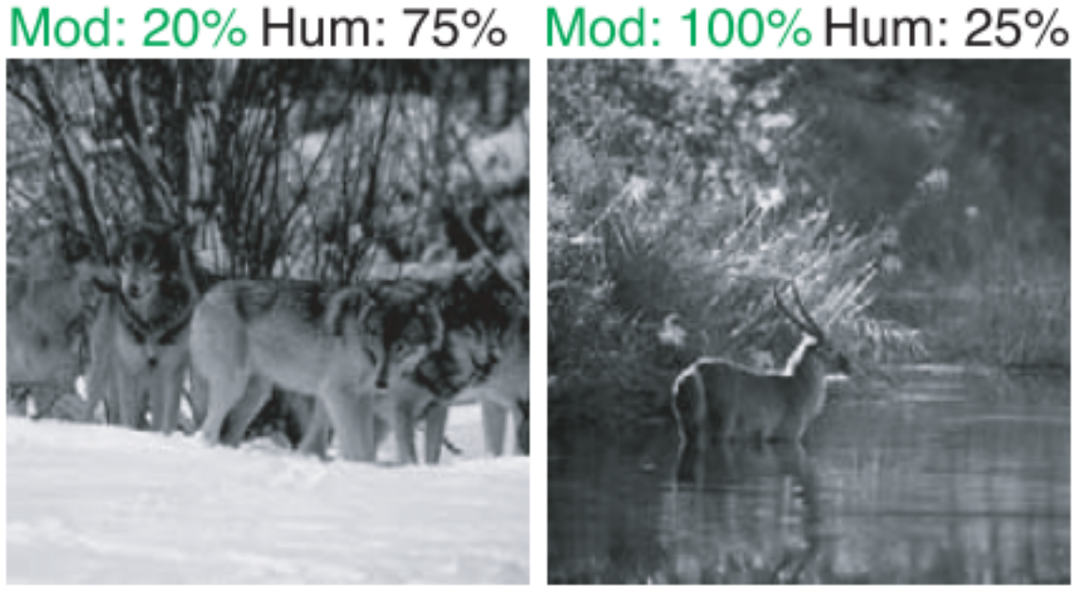
Serre et al. (2007)

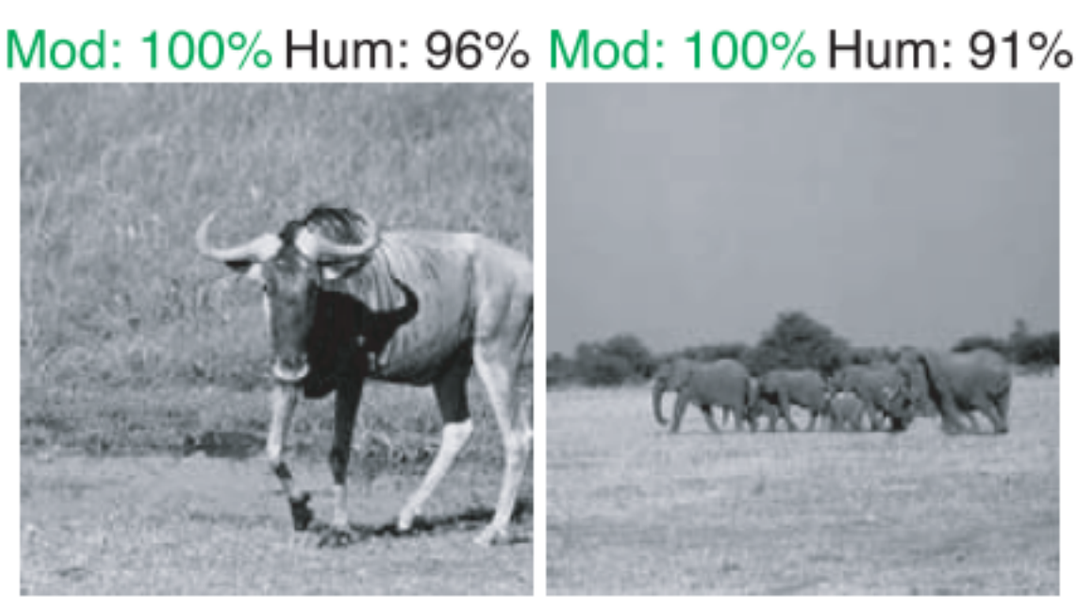

2.6
2.4
1.8
1.4
1.0
Performance (d')
Model
Human observers
Head
Close-body
Medium-body
Far-body

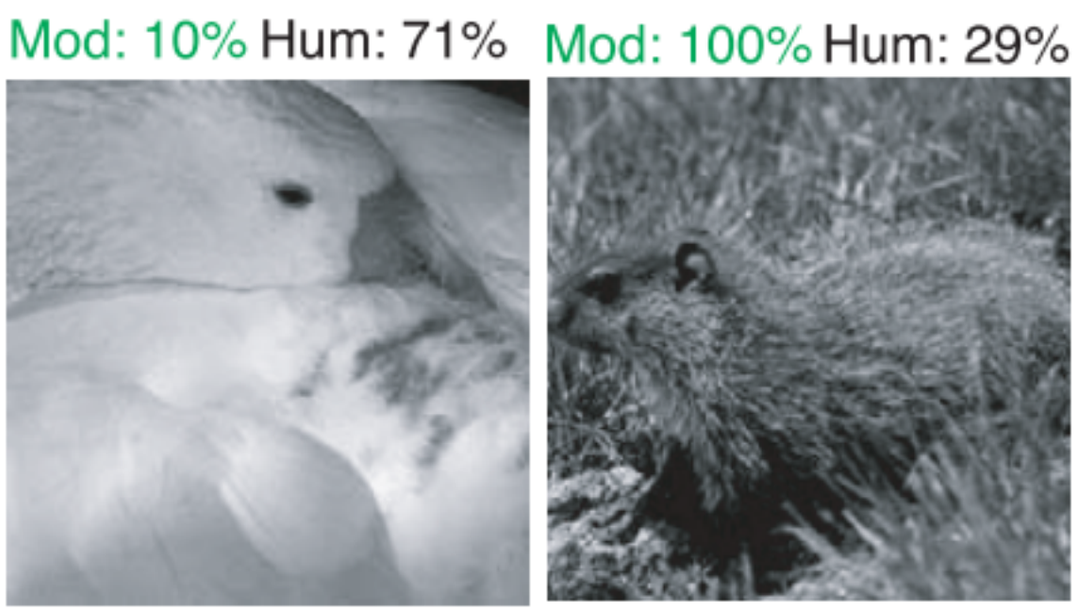
Feed-Forward Object Recognition Model

task: is there an animal in the picture?









Serre et al. (2007)

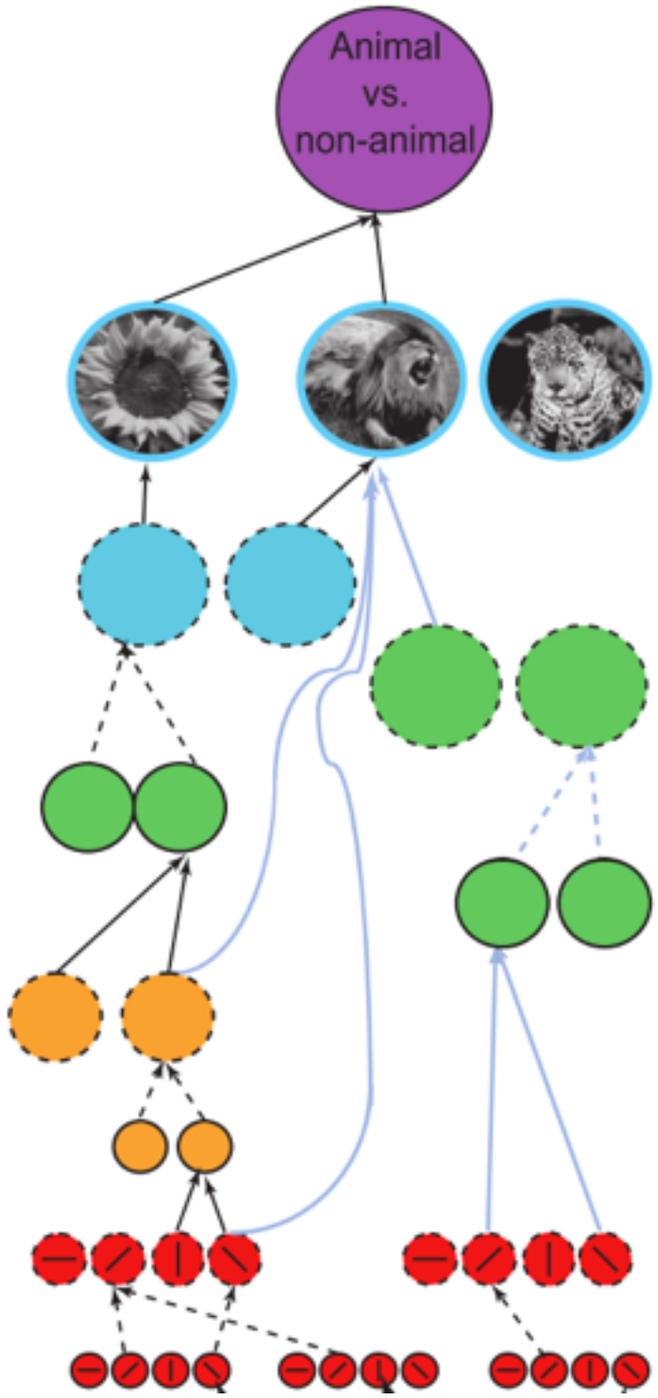
Feed-Forward Object Recognition Model

biologically-inspired processing pipeline

patches — receptive fields
building invariance
hierarchical processing

major drawback?

no feedback



Serre et al. (2007)

Discussion

Discussion

how closely should we aim to copy human vision?

is reverse-engineering human vision
a self-imposed limitation?

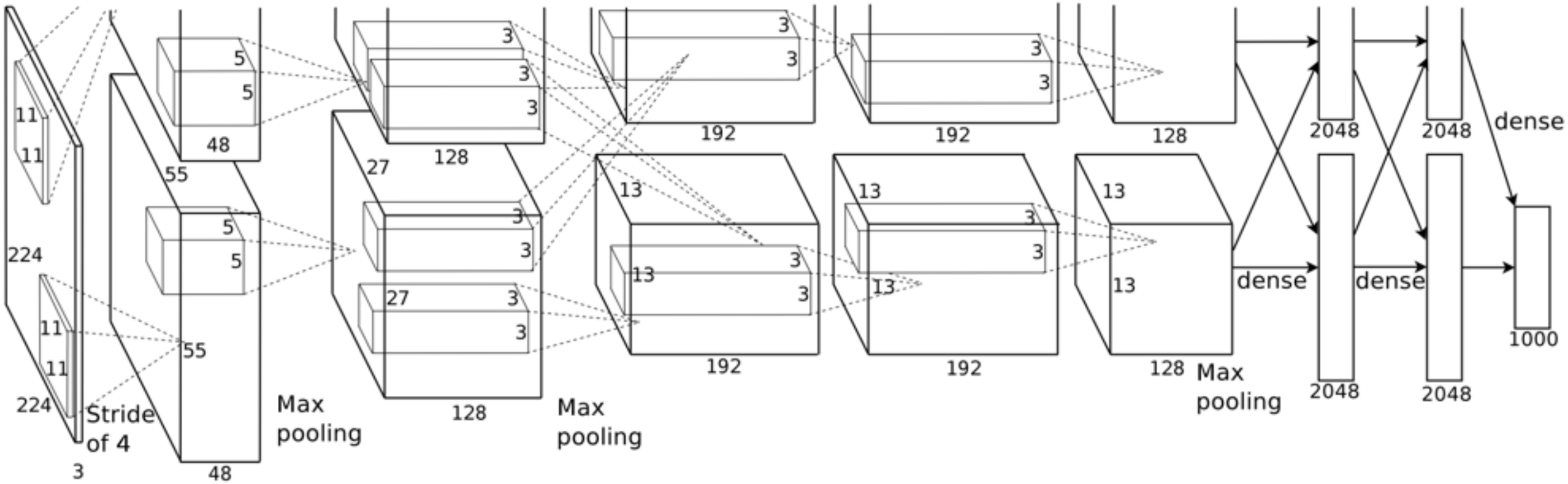
perfect recognition vs. visual understanding?

3. Modern Neural Networks

extremely large scale data and computation

Deep Convolutional Neural Network

650,000 cells — 60,000,000 parameters



input | convolution | pooling | … | convolution | pooling | fully-connected

Krizhevsky et al. (2013)

Deep Convolutional Neural Network

650,000 cells — 60,000,000 parameters

input | convolution | pooling | ... | convolution | pooling | fully-connected

Krizhevsky et al. (2013)

Deep Convolutional Neural Network

Want to know more about state-of-the-art neural networks?

CS 231N

http://vision.stanford.edu/cs231n/

Winter QT 2015

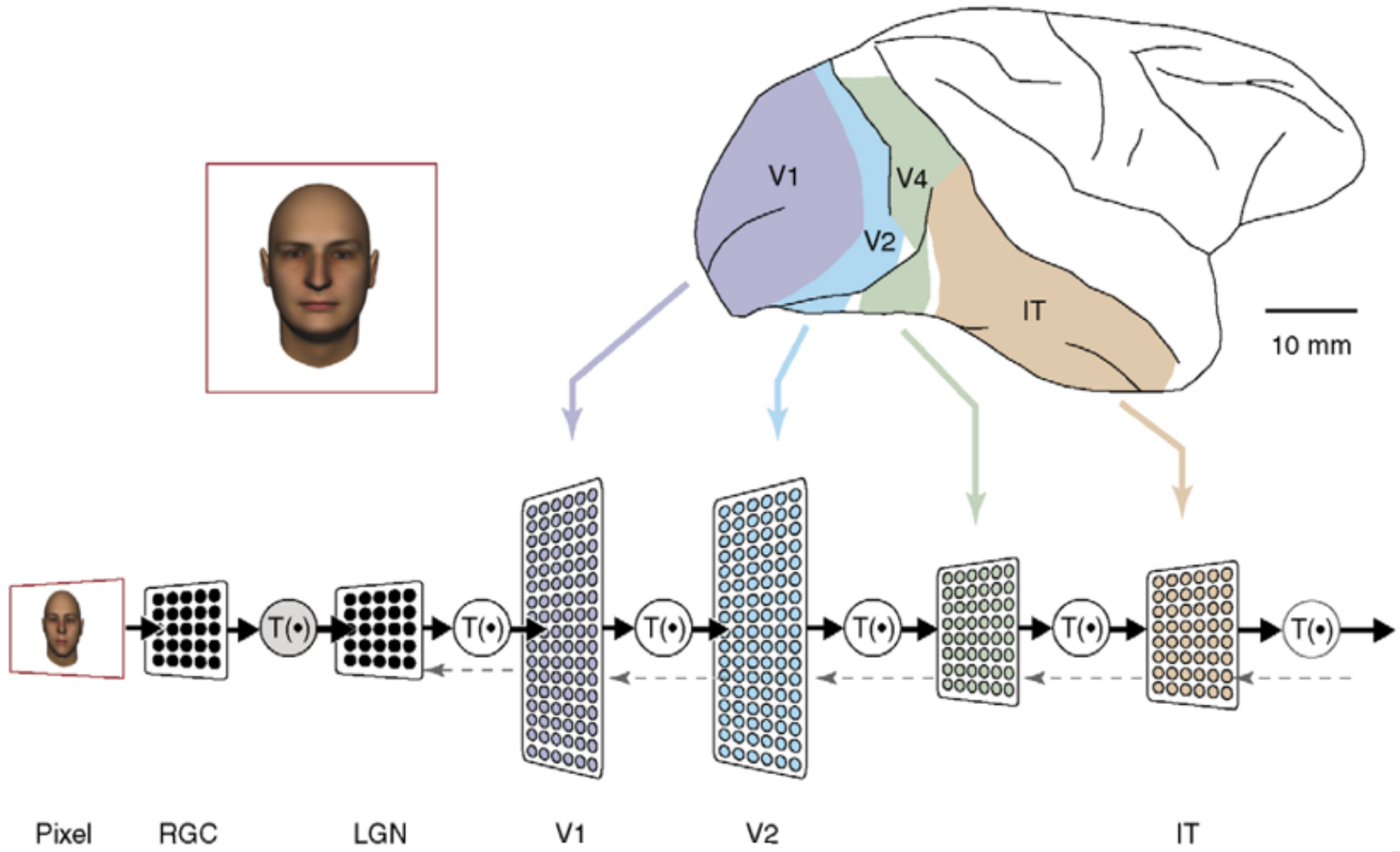
Fei-Fei Li & Andrej Karpathy

Object Recognition: Building Features and Invariance

visual processing is done in stages

each area performs a transformation on its inputs

invariance is built gradually across many successive steps



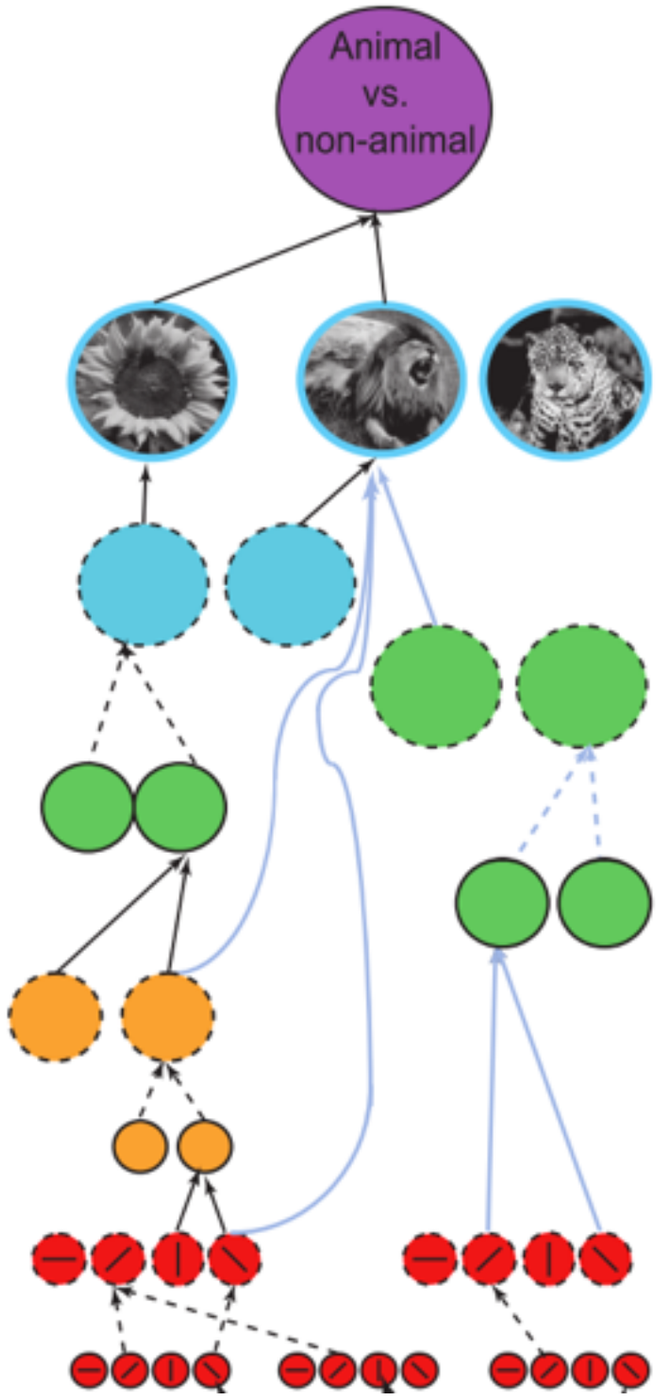
DiCarlo & Cox (2007)

Feed-Forward Object Recognition Model

biologically-inspired processing pipeline

patches — receptive fields
building invariance
hierarchical processing

major drawbacks?
no feedback
much less complex than human vision



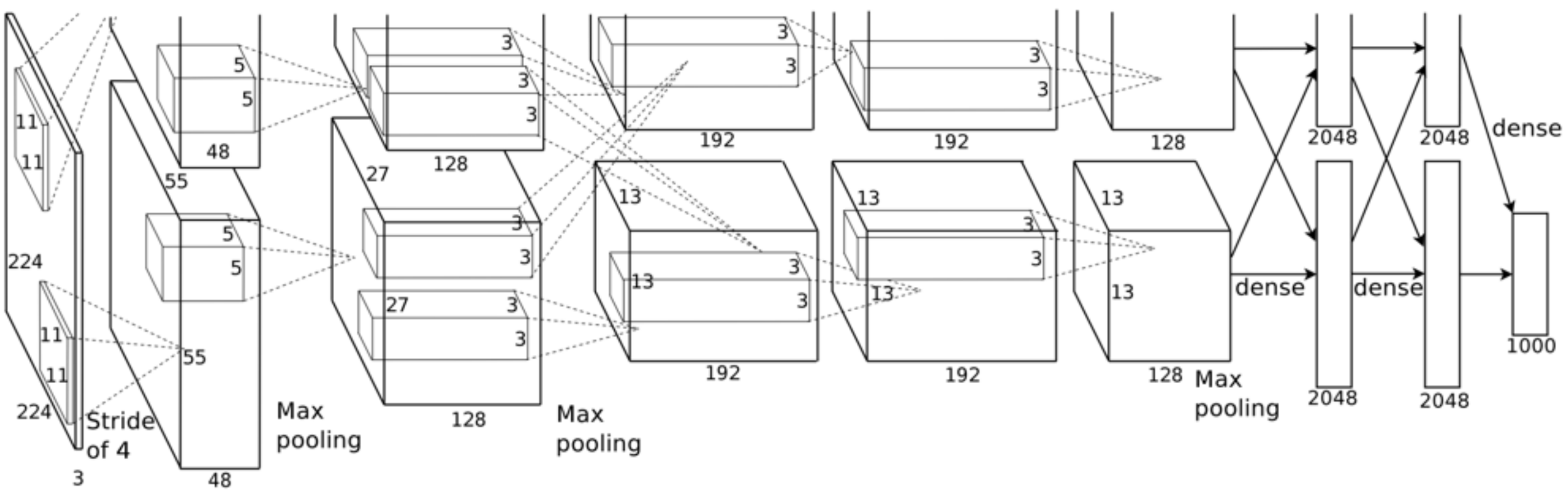
Serre et al. (2007)

Deep Convolutional Neural Network

extremely large scale data and computation

sudden, huge performance boost for recognition

if you want to know more, take CS 231N in Winter QT



Krizhevsky et al. (2013)

End-Quarter Feedback Forms