Object Recognition with Features Inspired by Visual Cortex

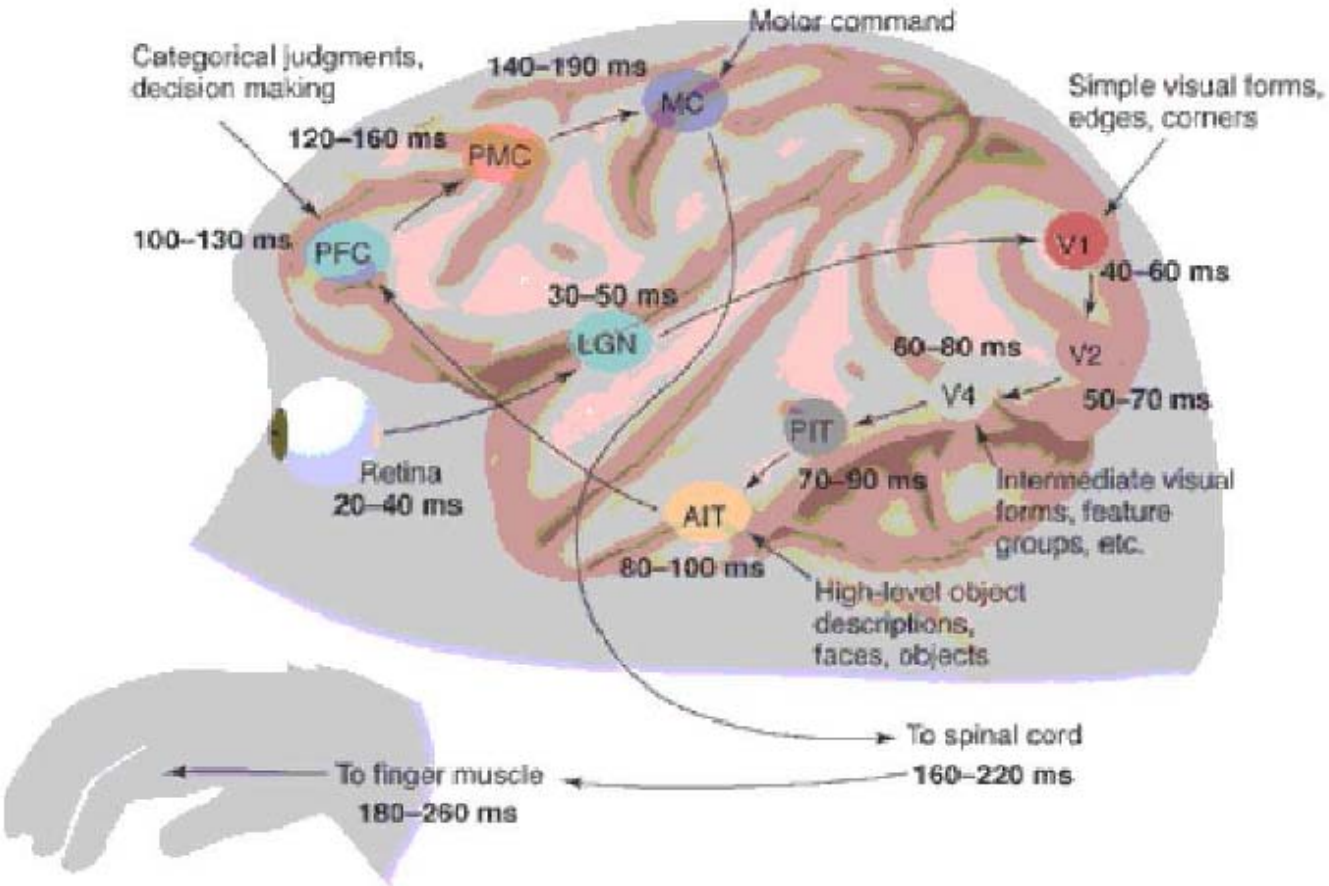
Serre, Wolf, Poggio
MIT 2005

Presented by Marius Buibas
10/17/2006

Summary

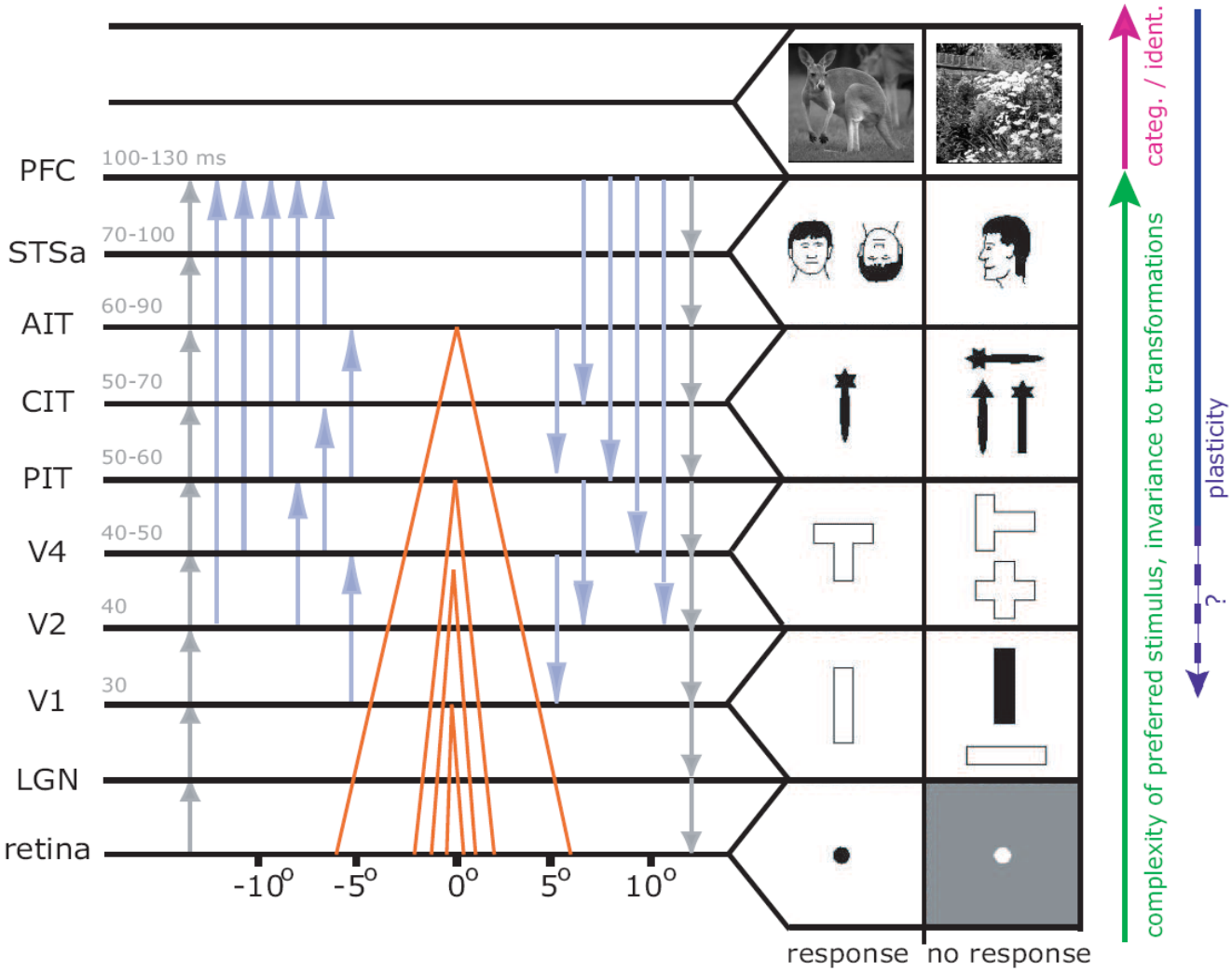
- Biology background
- Model Assumptions
- Algorithm Description
- Results
- Conclusions
- Critique/Forum/Questions

Brain Surgery



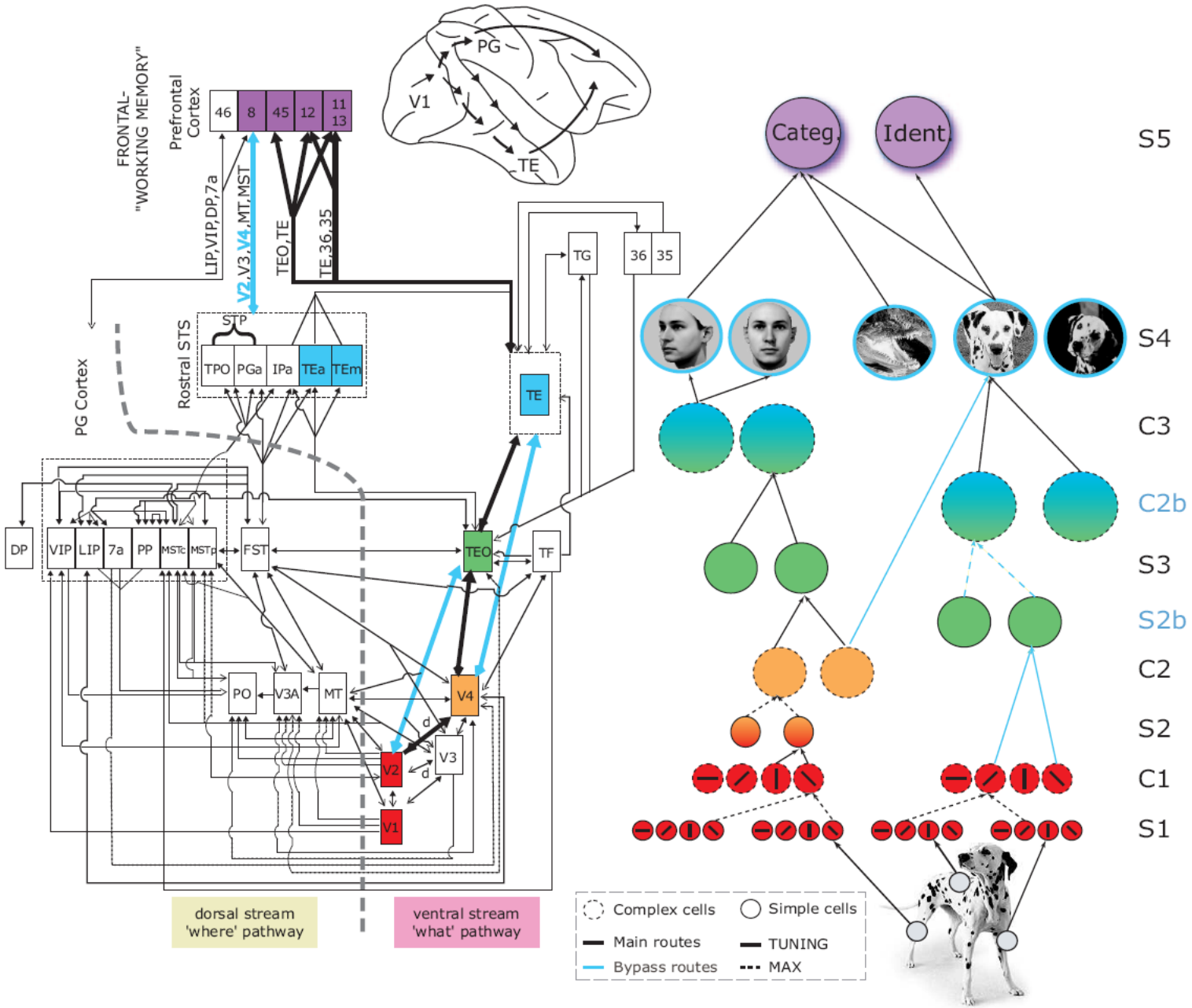
- Figure 1-3: The feedforward circuits involved in rapid categorization tasks. Numbers for each cortical stage corresponds to the shortest latencies observed and the more typical mean latencies [Nowak and Bullier, 1997; Thorpe and Fabre-Thorpe, 2001]. Modified from [Thorpe and Fabre-Thorpe, 2001].

Increased Selectivity

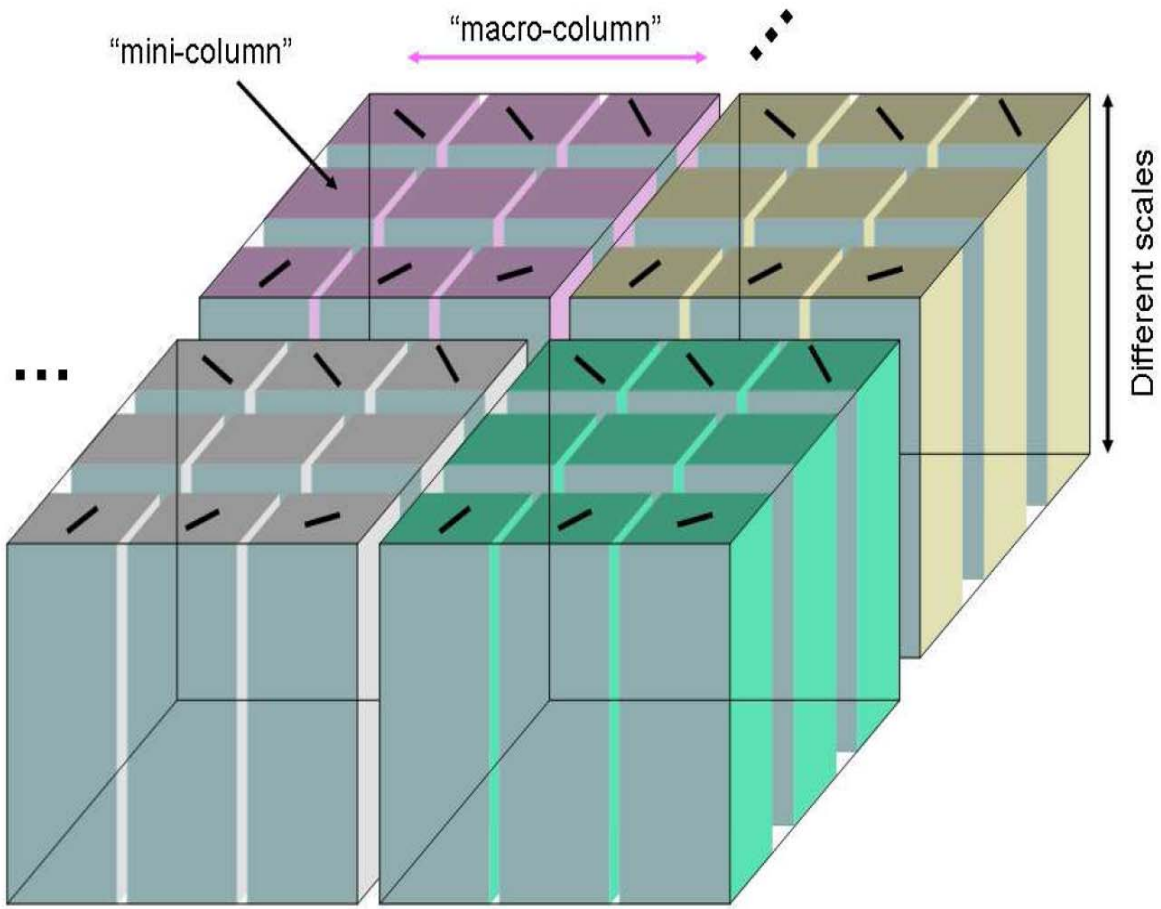


- Figure 1-2: The organization of visual cortex based on a core of knowledge that has been accumulated over the past 30 years. The figure is modified from[Oramand Perrett, 1994]mostly to include the likely involvement of prefrontal cortex during recognition tasks by setting task-specific circuits to read-out shape information from IT [Scalaidhe et al., 1999; Freedman et al., 2002, 2003; Hung et al., 2005].

System Architecture



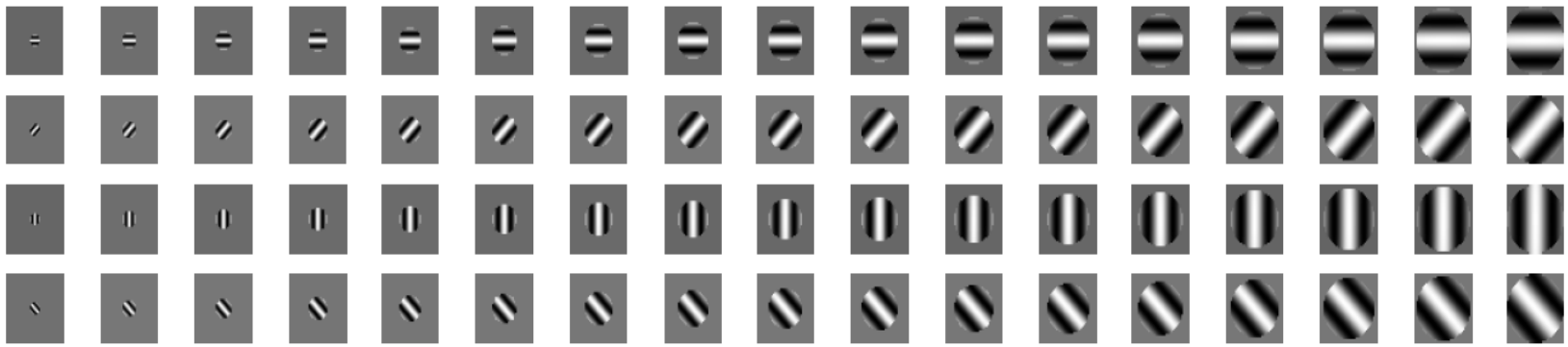
Functional Columnar Architecture



- Functional "columnar" organization in the model. Each basic mini-column contains a set of units all with the same selectivities, i.e., sharing the same weight vector w (e.g., a bar at a particular orientation at the S1 level) but different scales (e.g., 17 different scales/peak frequencies at the S1 level). Each portion of the visual field is analyzed by a macro-column which contains all types of mini-columns (e.g., 4 different orientations and 2 phases in the S1 case). The same organization is repeated in all layers of the model with increasingly complex and invariant units. Also note that there is a high degree of overlap in the portions of the visual field covered by neighboring macro-columns. Importantly note that we refer to columns in the model as functional primitives by analogy to the organization of visual cortex. Whether or not such functional columns in the model correspond to structural columns in cortex is still an open question.

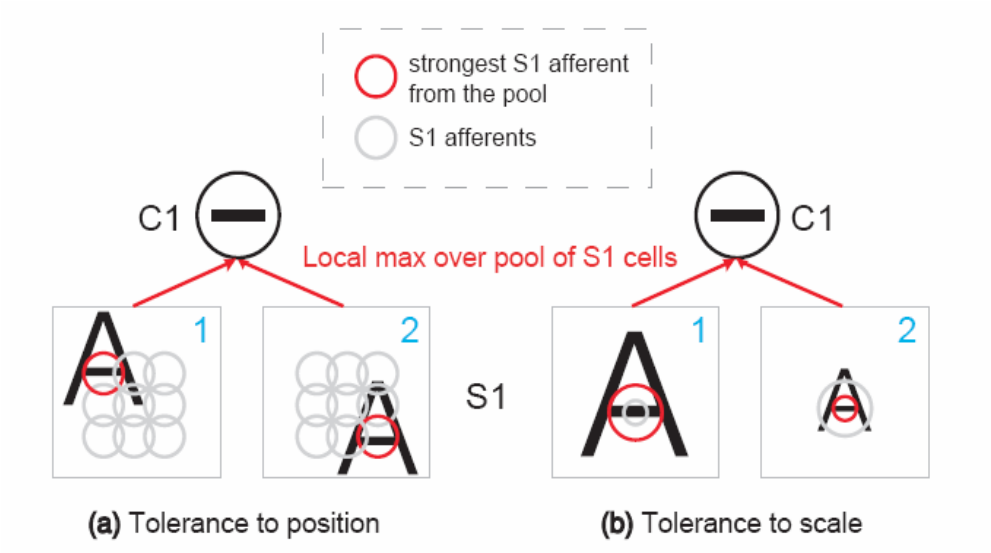
S1 Units – correspond to V1 Simple cells

- Bank of Gabor Filters analyze each pixel location in image at different phases, orientations or sizes
- Perform a TUNING operation with weighing vectors as Gabor filters



C1 units – Striate complex cells (Still V1)

- Each unit receives outputs of a group of S1 units from the first layer with same orientation (at both phases) and at slightly different positions and sizes/frequencies
- Response is a nonlinear MAX operation
- Provides increased tolerance to position and scale changes



S2 Units – In V2

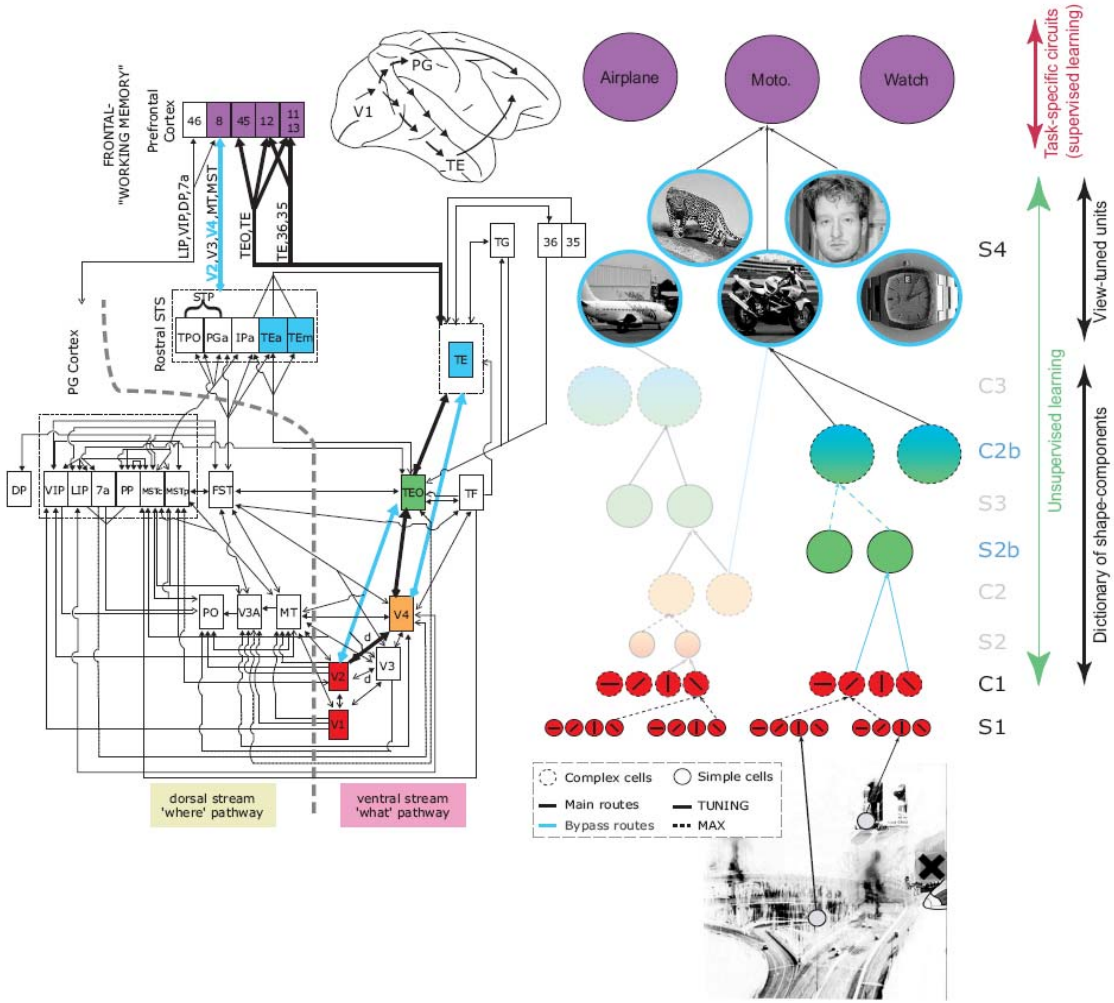
- Tune over C1 units at different preferred orientations over a small neighborhood.
- Both selectivity and complexity of preferred stimuli increased
- Tuning is learned in unsupervised manner from natural image
- ~1000 types of S2 units corresponding to different combinations of C1 units
- Layer organized in columns such that a small part of visual field is analyzed by one column containing all unity types at all scales (8 different scales from 8 C1 scales)

C2 Units – in layer 4 of V4

- Units pool over S2 units that are tuned to the same preferred stimulus (same vector W), but at different positions and scales.
- Selective for same stimulus as their afferents S2 units, but less sensitive to the position and scale of stimulus within receptive field

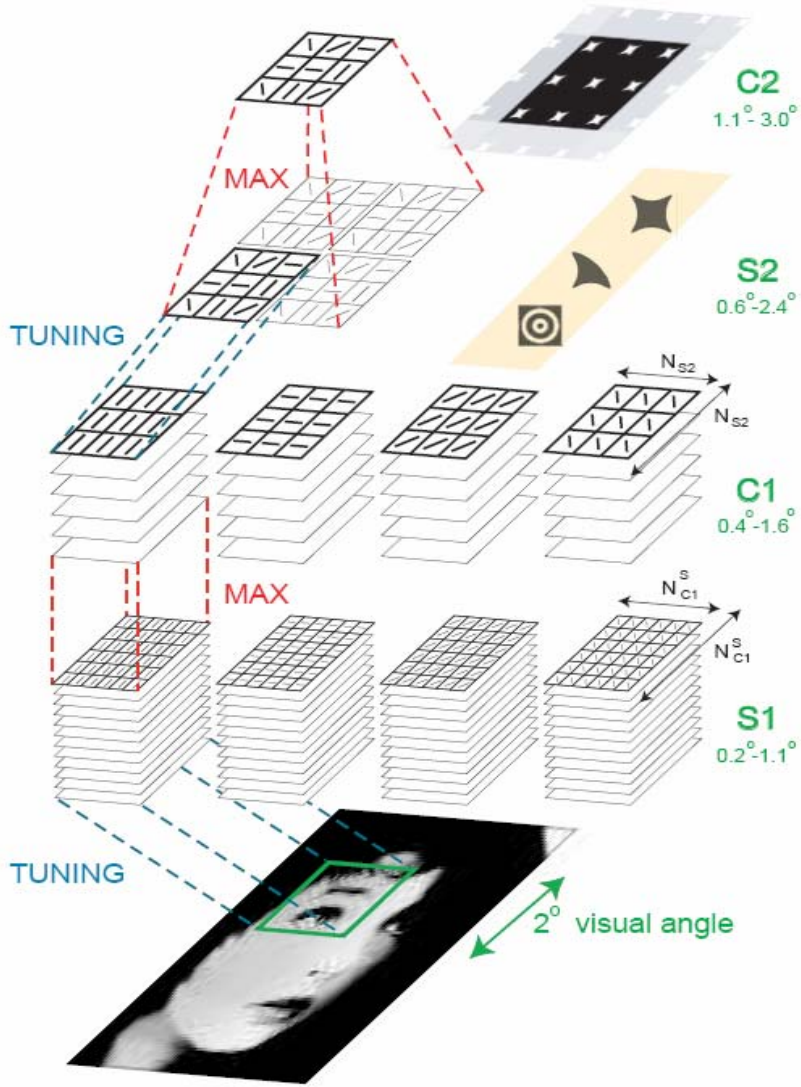
S3/C3 – in TEO – same process as S2/C2

- Input of ~100 C2 units
- Wider perceptive field
- Onto S4, in AIT



Algorithm Summary

- S1: Gabor Filters, 4 orientations 16 scales = 64 maps in 8 bands
- C1: Max over scales and positions for each band for an 8x8 grid
- Tune training patches at various sizes and all orientations, selected at random from C1 maps
- S2: Tune response Y, using Gaussian
- Max of S2 over all positions and scales, to obtain scale invariant C2 features



Benchmarking

- 50 positives and 50 negatives in training set, images at 140 pixels

Datasets		Bench.	C2 features	
			boost	SVM
Leaves (Calt.)	[24]	84.0	**97.0**	95.9
Cars (Calt.)	[4]	84.8	99.7	**99.8**
Faces (Calt.)	[4]	96.4	**98.2**	98.1
Airplanes (Calt.)	[4]	94.0	**96.7**	94.9
Moto. (Calt.)	[4]	95.0	**98.0**	97.4
Faces (MIT)	[7]	90.4	**95.9**	95.3
Cars (MIT)	[11]	75.4	**95.1**	93.3

Results

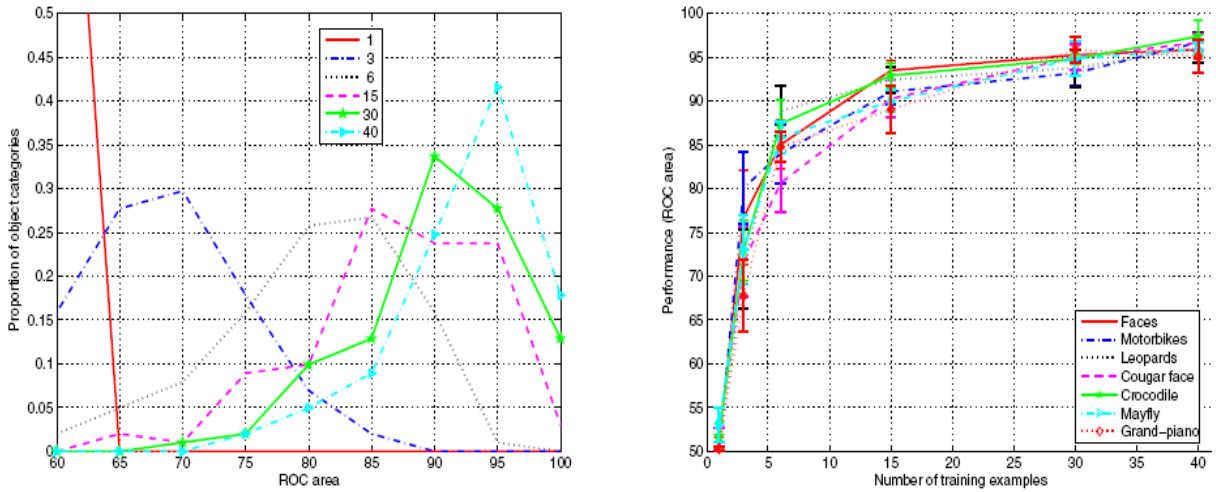


Figure 4. C2 features performance on the 101-object database for different numbers of positive training examples: (left) histogram across the 101 categories and (right) performance on sample categories, see accompanying text.

- Performance histograms for # of positive training examples for all 101 objects and by category
- Comparisons with SIFT
- SVM with Ada Boost classifier

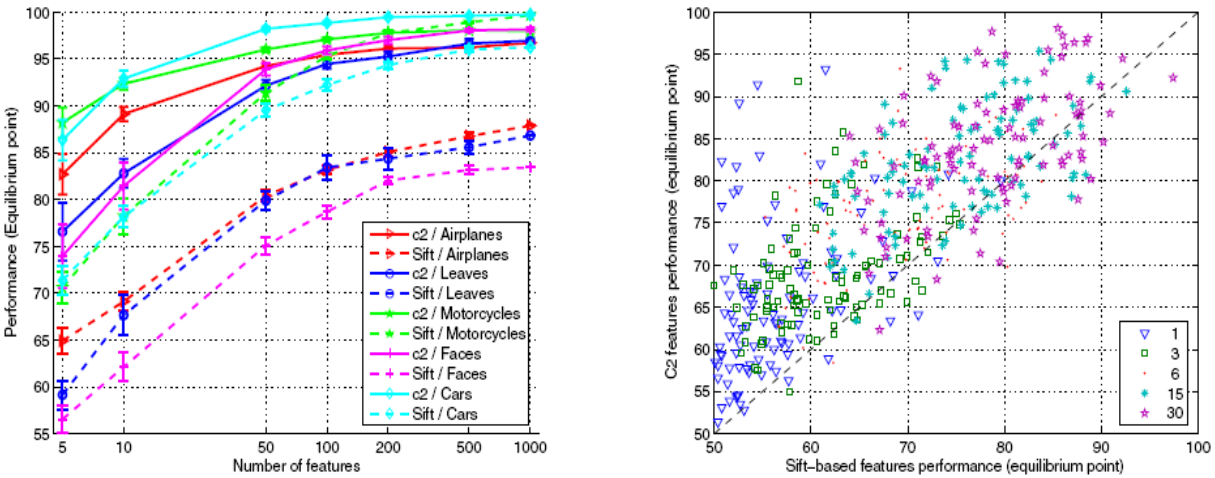


Figure 5. Superiority of the C2 vs. SIFT-based features on the Caltech datasets for different number of features (left) and on the 101-object database for different number of training examples(right).

Authors' Conclusions

- Biologically-motivated model outperforms more complex computer vision systems, with simple operations: template matching and max pooling
- Strength from built-in gradual shift- and scale-tolerance, that mimics visual cortical processing
- Hierarchical structure ease recognition problem by decomposing task into simpler ones at each layer
- Small number of features (50) required to achieve good error rates.

Critique/Forum/Questions