Learning Hierarchical Representations of Object Categories

Sanja Fidler and Aleš Leonardis
{sanja.fidler,ales.leonardis}@fri.uni-lj.si

University of Ljubljana
Faculty of Computer and Information Science
Visual Cognitive Systems Laboratory

Outline

- Motivation
- Related work
- Our hierarchical representation
- Results
- Summary

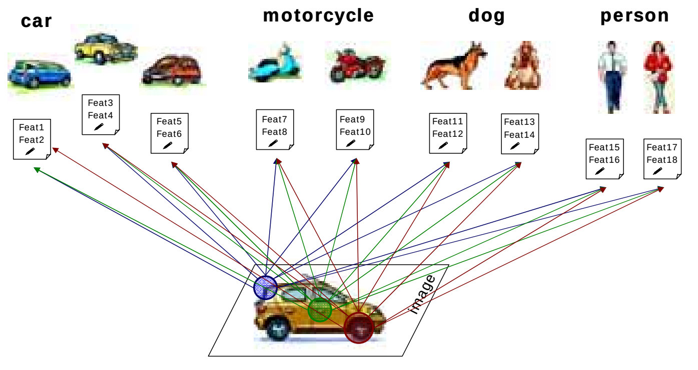
Motivation

- Challenges in cognitive systems (ECVision Roadmap)
  - **Representation**
    - Representation enabling to deal with approx. 30.000 visual categories
    - Representation that would enable connection with languange, manipulation, affordances
  - **Learning**
    - Dealing with a large number of categories, the importance of learning becomes more pronounced
    - Various representations cannot be put into the system by hand
    - Adaptation to ever changing environment
  - **Detection, categorization**
    - Fast categorization and detection of multiple objects on multiple scales in an image

Requirements

- Requirements for a general visual system capable of recognizing **a large number** of object categories:
  - **Hierarchical representation** [Tsotsos, Geman, Rolls]

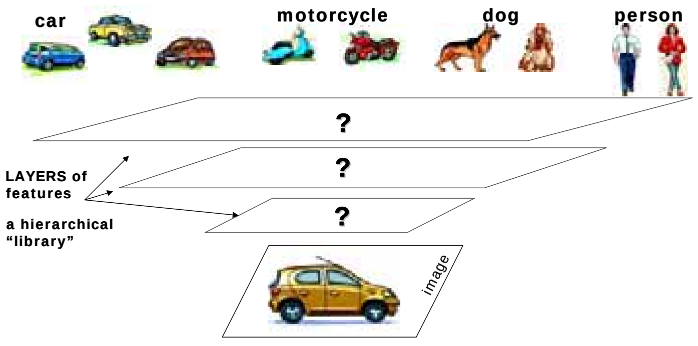
Limitations of flat representations



Matching must be **computationally feasible**.

Hierarchical representation

Features on layers should be manifested as **models- parts**.

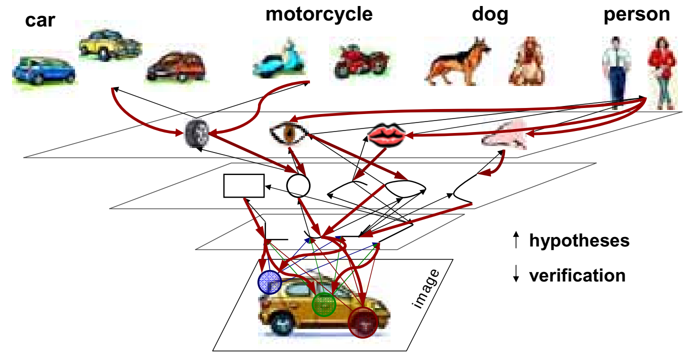


What features should the layers consist of? How many layers?

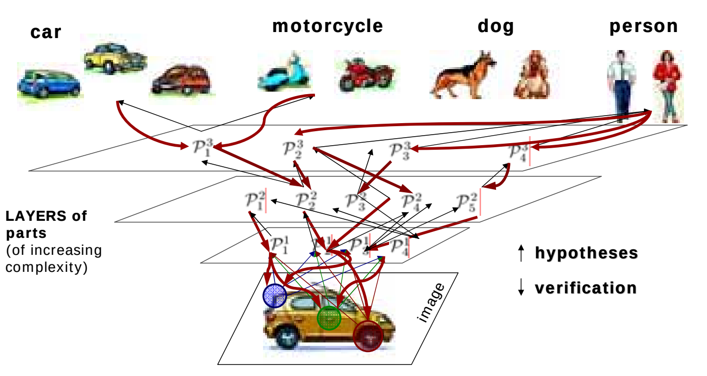
Requirements

- Requirements for a general visual system capable of recognizing **a large number** of object categories:
  - Hierarchical representation
  - **Computational plausibility** (compositionality [Geman], indexing & matching [Califano, Amit])

Detection: Indexing and matching



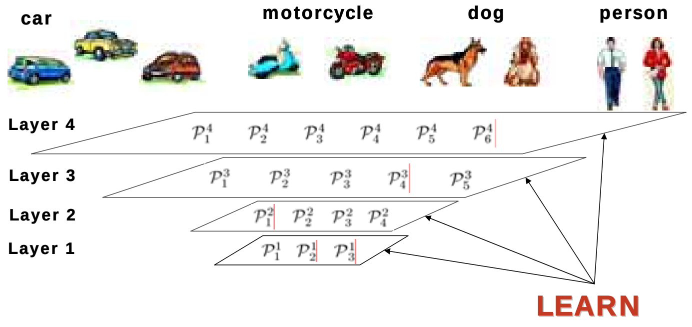
Detection: Indexing and verification



Requirements

- Requirements for a general visual system capable of recognizing **a large number** of object categories:
  - Hierarchical representation
  - Computational plausibility (compositionality, indexing & matching)
  - **Statistics driven learning** (unsupervised learning [Edelman, Mel, Barlow])

Bottom-up learning

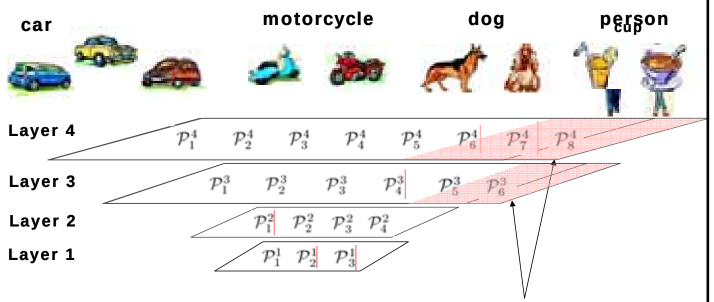


hierarchical library

Requirements

- Requirements for a general visual system capable of recognizing **a large number** of object categories:
  - Hierarchical representation
  - Computational plausibility (compositionality, indexing & matching)
  - Statistics driven learning (unsupervised learning)
  - **Fast, incremental (continuous) learning** [ECVision Roadmap]

Bottom-up learning



add parts without restructuring the complete hierarchy
(possibly add only to higher layers)

hierarchical library

Related work

- Hierarchical representation
  - Riesenhuber, Serre & Poggio (HMAX), Mutch & Lowe, Lecun, Amit & D. Geman, S. Geman, Torralba, Borenstein, Epstein & Ullman, Scalzo & Piater, Bouchard & Triggs
- Compositionality, indexing & matching
  - S. Geman, Amit & Geman
- Statistics driven learning (unsupervised learning)
  - Serre & Poggio, Fleuret & D. Geman, Scalzo & Piater, Lecun
- Fast, incremental (continuous) learning
  - Opelt, Lecun

⇨ Our work: **all issues in unified novel framework** in the pursuit of a general categorization system

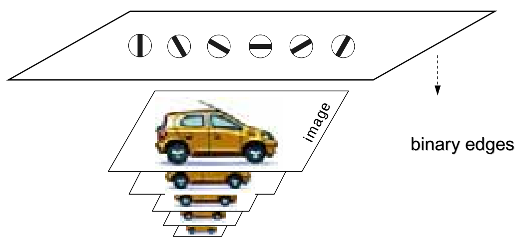
Our hierarchical architecture

- Starts with simple, local features and **learns** more and more complex **compositions**
- Learns layer after layer to exploit the regularities in natural images as efficiently and compactly as possible
- Builds computationally feasible layers of parts by selecting only the most statistically significant compositions of specific granularity
- Learns lower layers in a category independent way (to obtain optimally sharable parts) and category specific higher layers which contain only a small number of highly generalizable parts for each category
- New categories can efficiently and continuously be added to the representation without the need to restructure the complete hierarchy
- Implements parts in a robust, layered interplay of indexing & matching

Learning a hierarchical representation

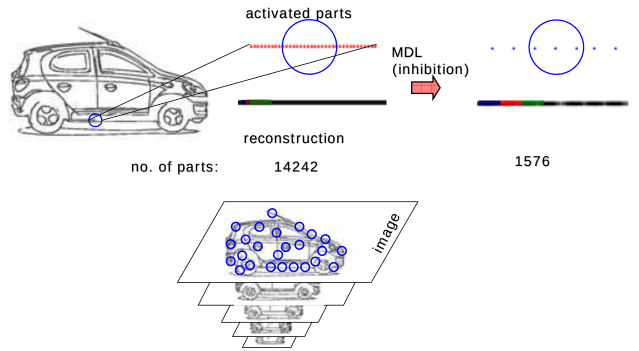
- Layer 1 consists of oriented filters (simple – only a very small number of features describe most of the image)
- Applied in all image pixels, not only around interest points
- Must account for multiple scales

Layer 1
(oriented
filters)



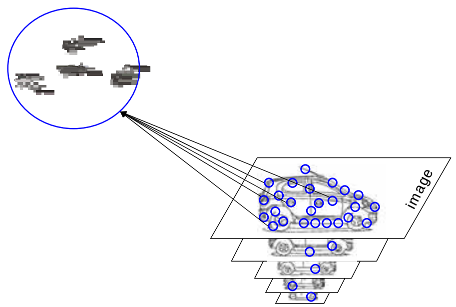
Learning a hierarchical representation

- Reduce redundancy to allow for computationally feasible learning (inhibition)
- Learn combinations in local neighborhoods of each image part



Learning a hierarchical representation

- Reduce redundancy to allow for computationally feasible learning (inhibition)
- Learn combinations in local neighborhoods of each image part

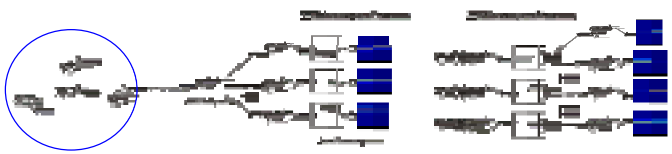


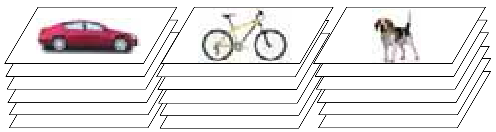
Acquire statistics for:

- each image part
- each scale

Learning a hierarchical representation

- Reduce redundancy to allow for computationally feasible learning (inhibition)
- Learn combinations in local neighborhoods of each image part



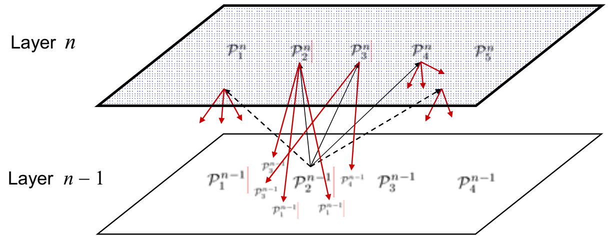


Acquire statistics for:

- each image part
- each scale
- a number of images

Selection of parts: Complexity issues

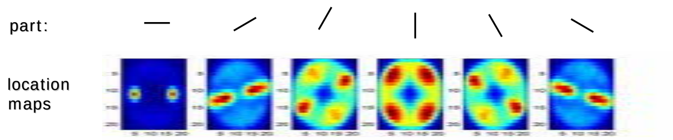
- **Hierarchical indexing and verification**



Complexity of verification: parts should **have a small number of components**, number of indices to higher level parts should be relatively low (especially for lower level parts, which appear more frequently)

2nd Layer

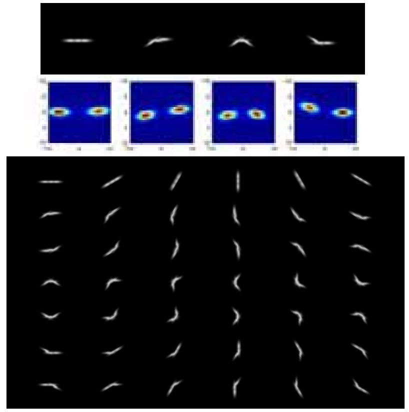
Results for one part conditional to the central part (normalized to orientation 0):



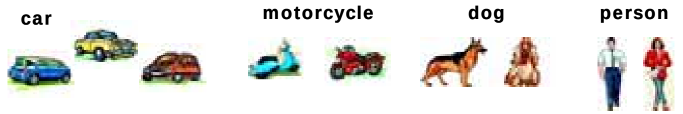
(for 7000 images)

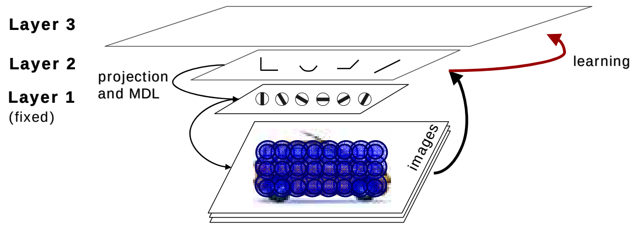
2nd Layer

- Final Layer 2 parts with maps

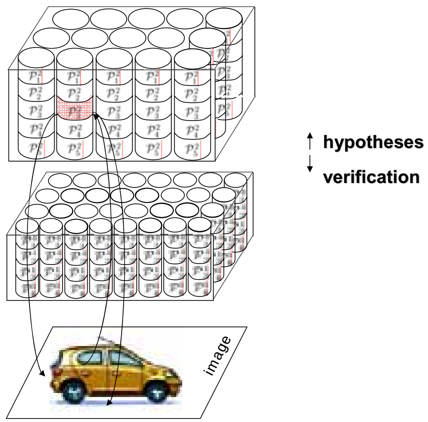


Towards 3rd Layer





Detection: Indexing and verification



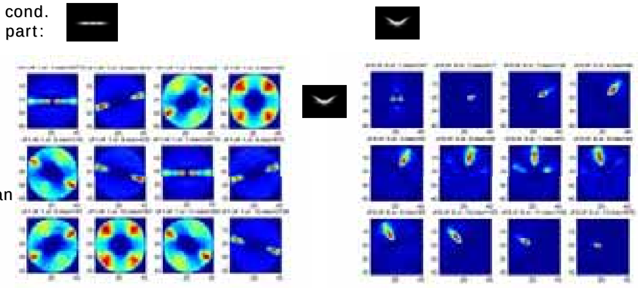
3rd Layer

Results for one part conditional to the central part (normalized to orientation 0):

cond. part:

central part:

gaussian maps



Results

- 3rd Layer – obtained on 2,000 clip arts

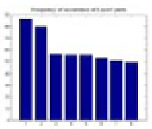
Layer 3

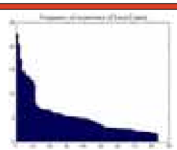
- Statistical confirmation to Gestalt laws (colinearity, parallelism, circularity)

Results – 15 categories (3000 images)

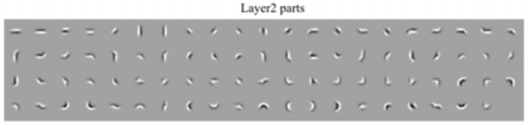
- Learned parts

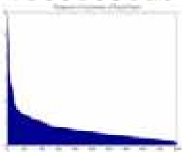
Layer1 parts



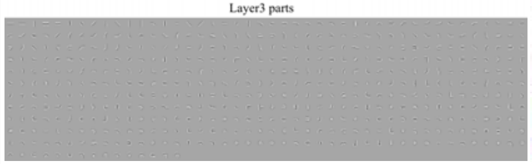


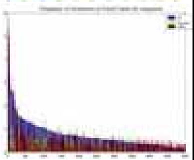
Layer2 parts





Layer3 parts

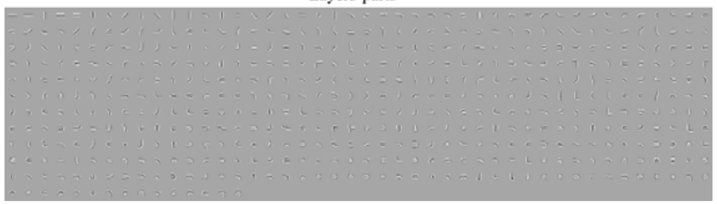




Results, Layer 3

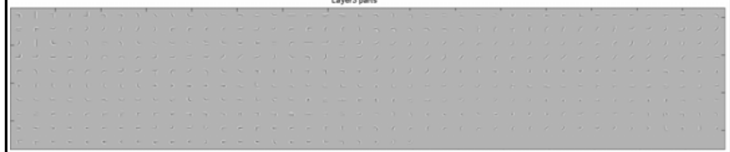
- All

Layer3 parts



- Mugs

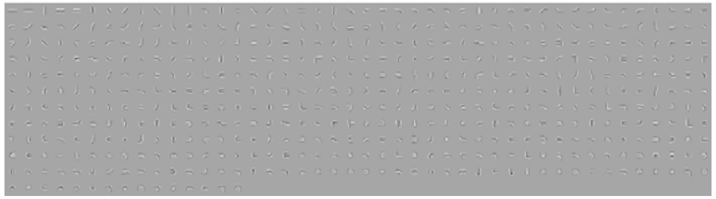
Layer3 parts



Results, Layer 3

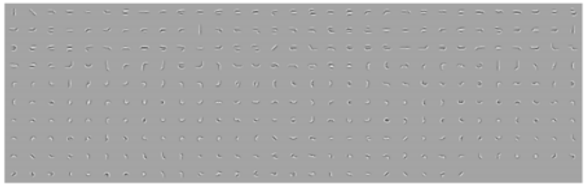
- All

Layer3 parts



- Faces

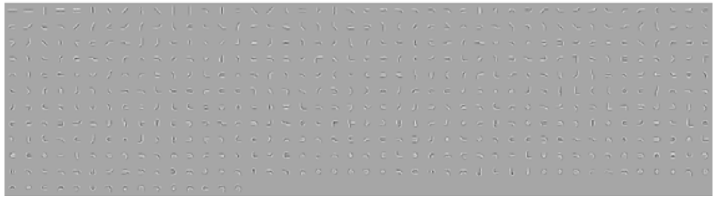
Layer3 parts



Results, Layer 3

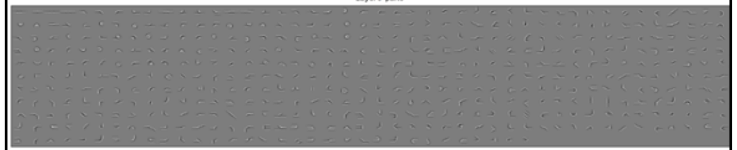
- All

Layer3 parts

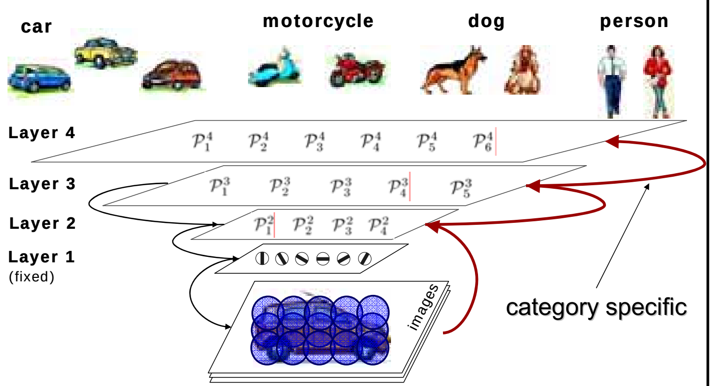


- Cars

Layer3 parts

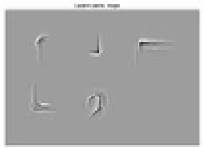


Bottom-up learning



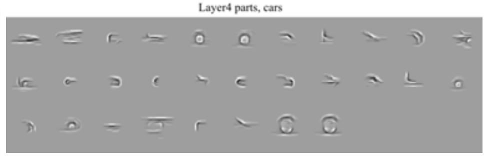
Results, Layer 4

- Layer 4 (category specific)



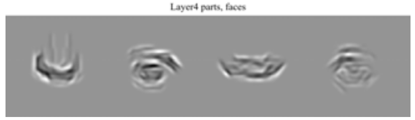
mugs

Layer4 parts, cars



cars

Layer4 parts, faces



faces

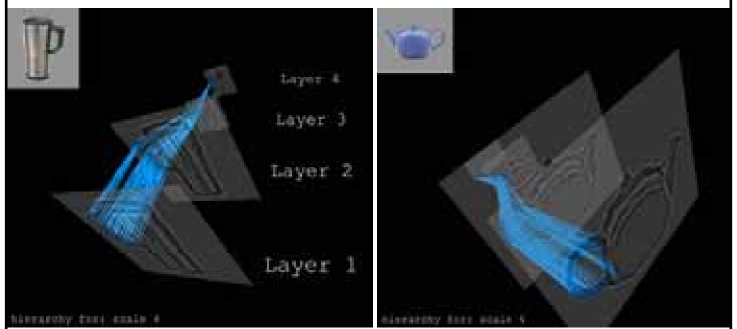
Specific categories - mugs

Sanja Fidler and Aleš Leonardis. *Learning Hierarchical Representations of Object Categories*. EU Cognition meeting, Munich 2007 42

Results - Specific categories

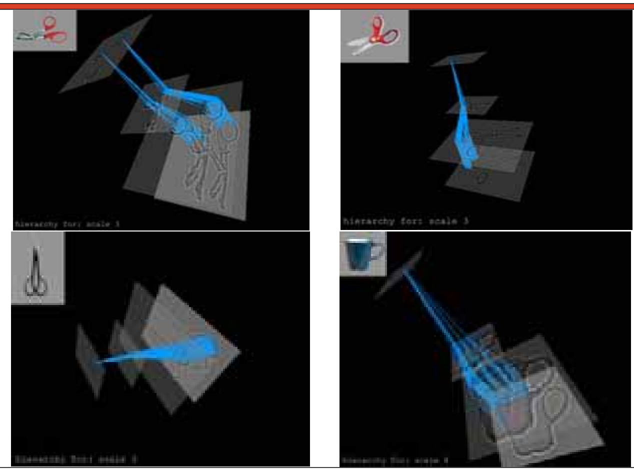
Sanja Fidler and Aleš Leonardis. *Learning Hierarchical Representations of Object Categories*. EU Cognition meeting, Munich 2007 43

Results - Specific categories



Sanja Fidler and Aleš Leonardis. *Learning Hierarchical Representations of Object Categories*. EU Cognition meeting, Munich 2007 44

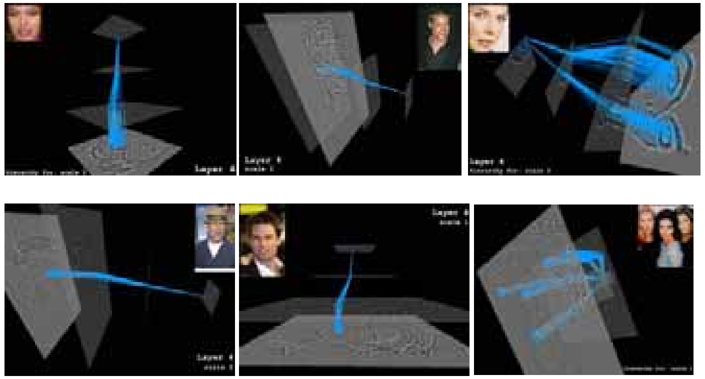
Results - Specific categories



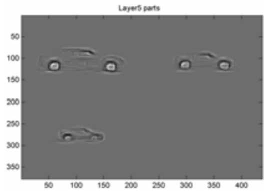
Sanja Fidler and Aleš Leonardis. *Learning Hierarchical Representations of Object Categories*. EU Cognition meeting, Munich 2007 45

Results - Specific categories, faces

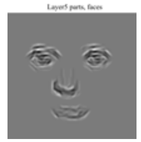
- Detection of Layer4 parts



Results – Layer 5

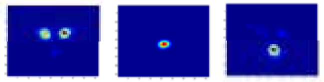


cars



faces

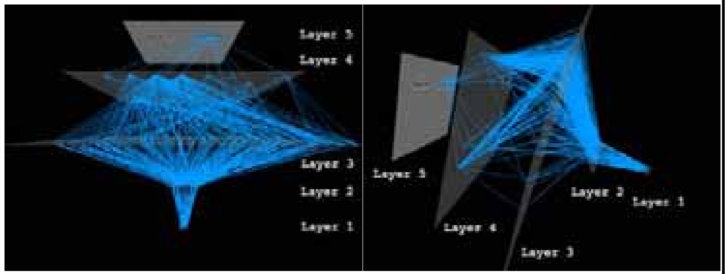
mugs



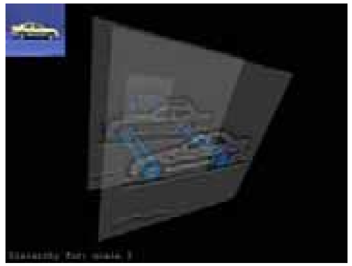
learned maps for faces

Results - Specific categories, cars

- Learned compositionality for cars



Results - Specific categories, cars

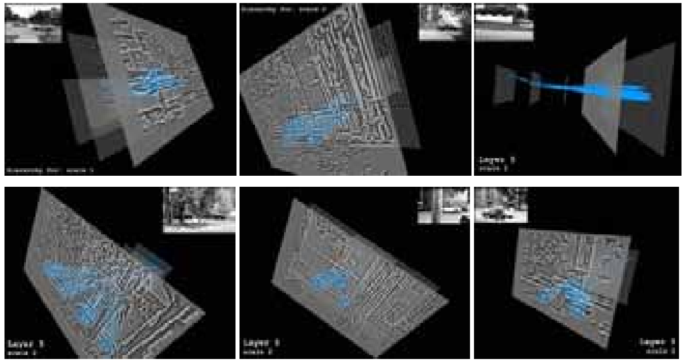


Layer 4

Layer 5

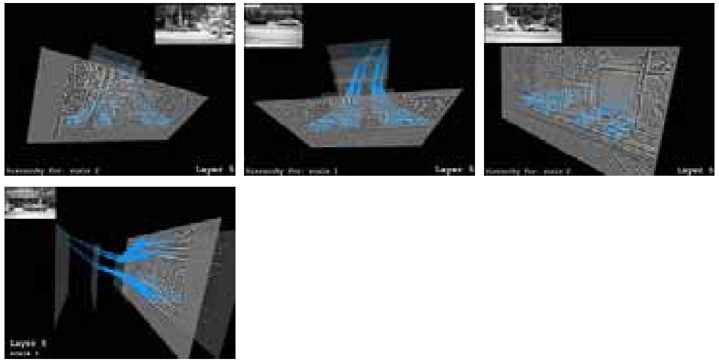
Results - Specific categories, cars

- Detection with Layer 5



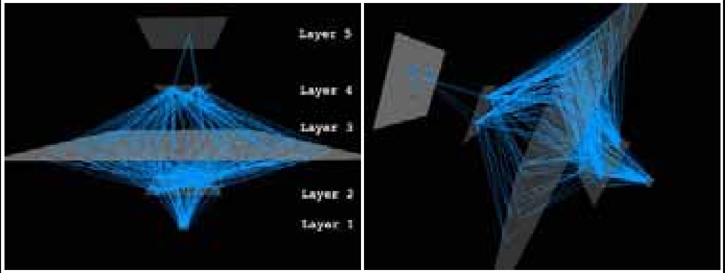
Results - Specific categories, cars

- Detection with Layer 5



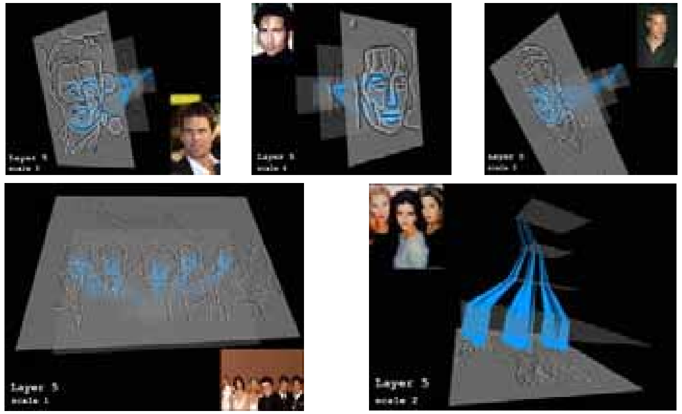
Results - Specific categories, faces

- Learned compositionality for faces



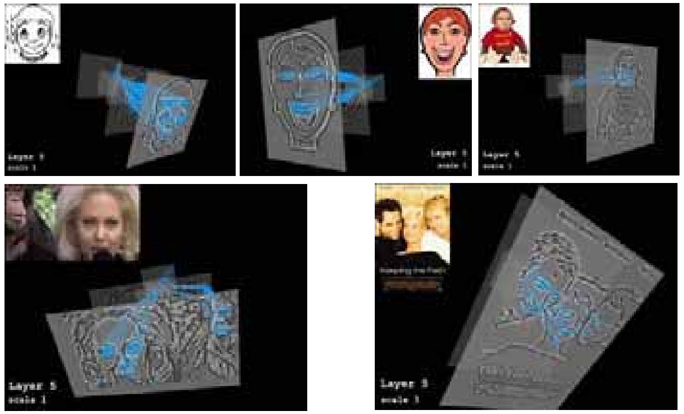
Results - Specific categories, faces

- Detection of Layer5 parts



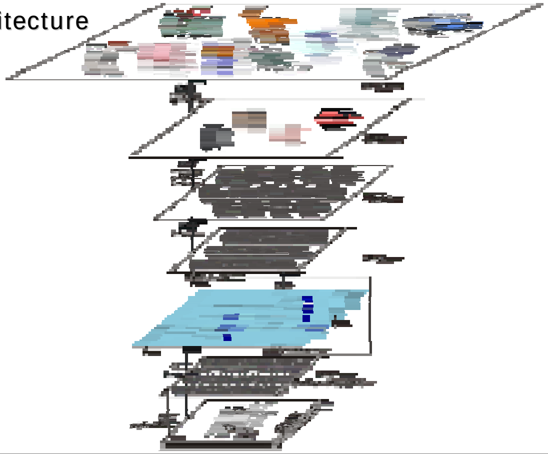
Results - Specific categories, faces

- Detection of Layer5 parts

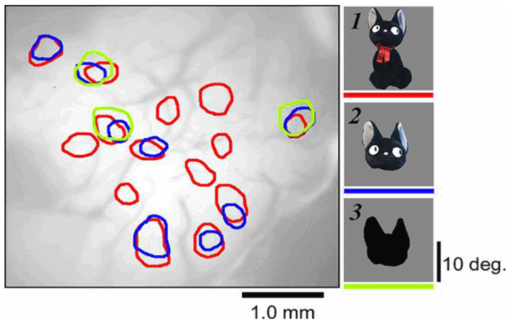


Overview of the architecture

- Architecture



Distributed local representations



M. Tanifuji, Nature 2001

Summary

- **Hierarchical representation of visual structure**
  - Computational feasibility
  - Fast indexing and matching
  - Generalizations through spatial flexibility and grouping of similar parts throughout the hierarchy
- **Unsupervised statistical learning**
  - Category independent (layers 1-3)
  - Category specific (layers 4, 5)
  - Small number of sharable parts relative to varying complexity and specificity
  - Incremental learning

For details please see:

S. Fidler and A. Leonardis. *Towards Scalable Representations of Object Categories: Learning a Hierarchy of Parts.* To appear in CVPR 2007.

S. Fidler, G. Berginc, A. Leonardis. *Hierarchical Statistical Learning of Generic Parts of Object Structure.* CVPR 2006.