Learning Compositional Models

A.L. Yuille

Bloomberg Distinguished Professor

Depts. Cognitive Science and Computer Science

Johsn Hopkins University

Representation Learning For Computer Vision

- Learning currently plays a major role in Computer Vision. Deep Nets are the most obvious example.
- *They have many desirable properties. A hierarchical structure and, in particular, can be trained by differentiation, because their outputs are differentiable functions of their parameters/weights and of the input images.*
- This makes learning conceptually easy although technically complex. There are many recent papers which improve them – robust features, two-batch-normalization, weight-standardization, softened ReLu functions (only the few that I know about).
- ***But, for many reasons, I favor hierarchical networks that are more semantic and explainable than Deep Networks. We need networks that can perform many visual tasks with the same underlying representation. We need the ability to perform domain transfer and domain generalization.***
- *How can we learn neural architectures which can do this? Differentiation and backprop (AKA chain rule of differentiation) are not enough.*

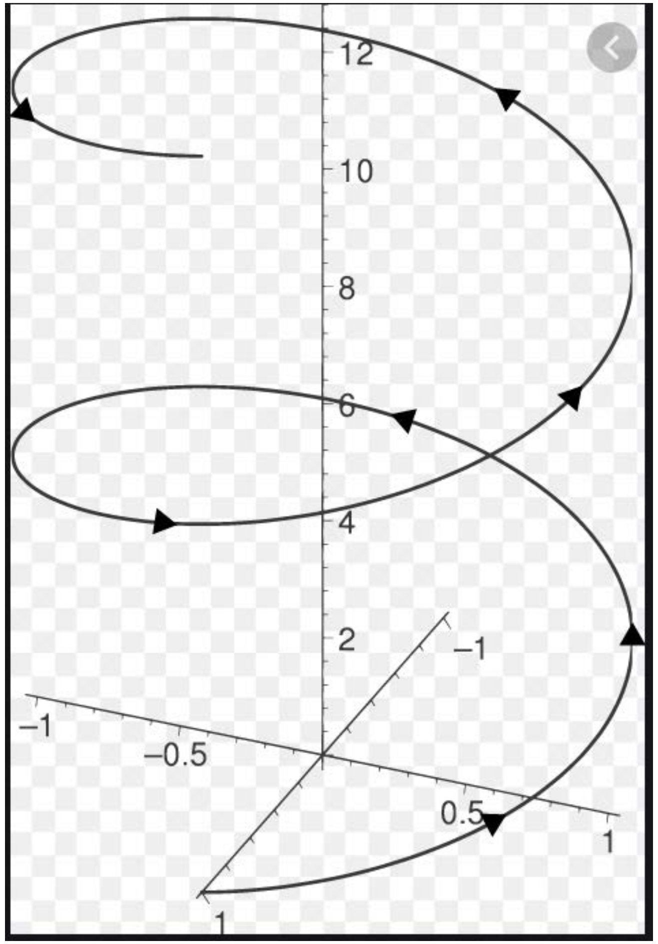
The Helix of Progress of Computer Vision.

- *“Those who forget history are condemned to repeat it” George Santayana.*

- ***This is true in machine learning and computer vision. There are fashions which keep repeating and older topics get reinvented, renamed, and rebranded.***

- *We should not forget the work before Deep Nets. It will surely come back into fashion.*

- Is Computer vision repeatedly going around in a circle?
- No, because each time we go round the circle our technical skills and our performance increases. There is a “helix of progress” (***Andrew Blake – personal communication).***



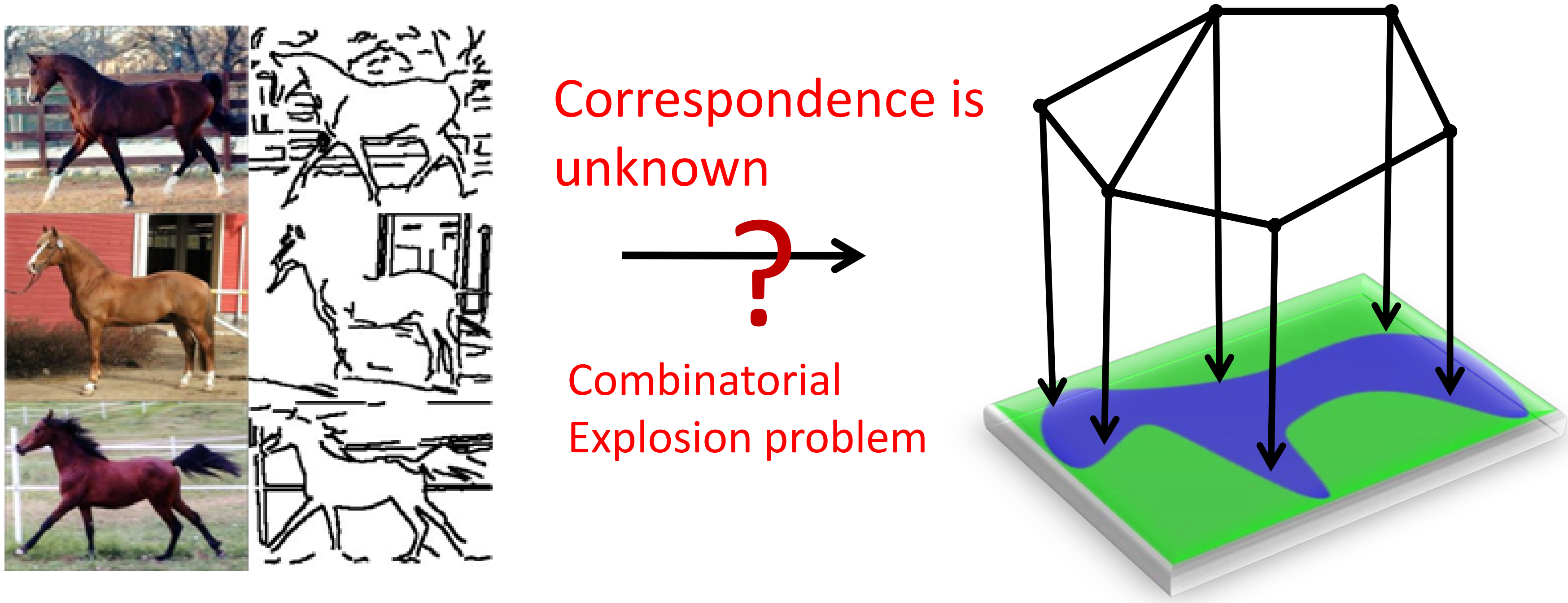
Part 1: Unsupervised Structure Learning

- The goal is to learn hierarchical semantic representations. Compositional, in the sense that parts are composed of subparts (not compositional mathematical functions, as in deep networks).
- This relates to the literature on hierarchical probabilistic grammars. These include AND/OR graphs (e.g., Zhu & Mumford 2007).
- Start with a dictionary of basis elements, e.g., oriented edges, and hierarchically cluster to get higher level dictionaries of parts and subparts. Intuition: learn by identifying *suspicious coincidences* and preforming *competitive exclusion*.

Leo Zhu et al. (2008, 2010). See also Fidler and Leonardis (2007).

Unsupervised Hierarchical Structure Learning

- Task: given 10 training images, *no labeling, no alignment, highly ambiguous features*.
  - Estimate Graph structure (nodes and edges)
  - Estimate the parameters.

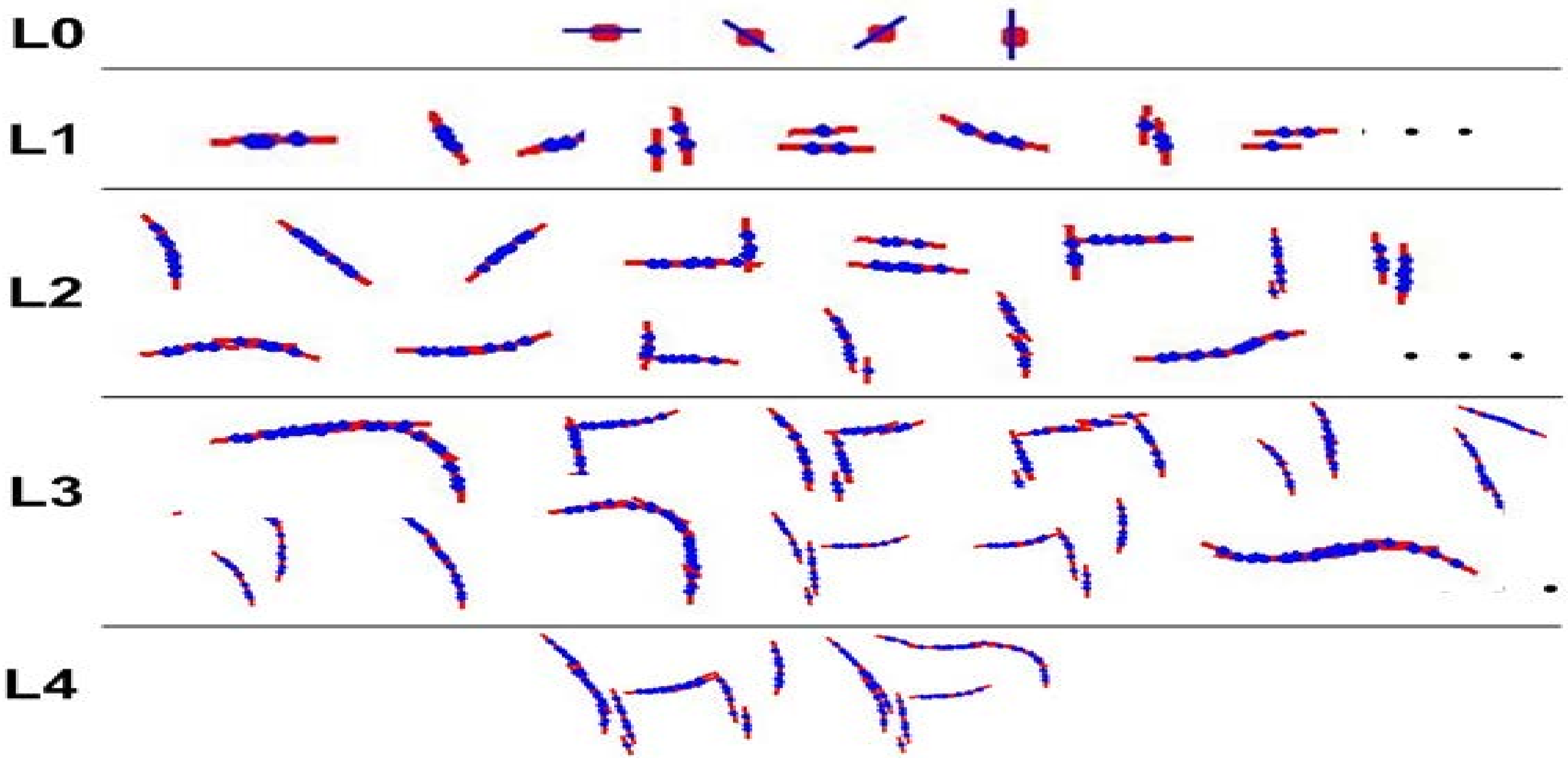


Correspondence is unknown

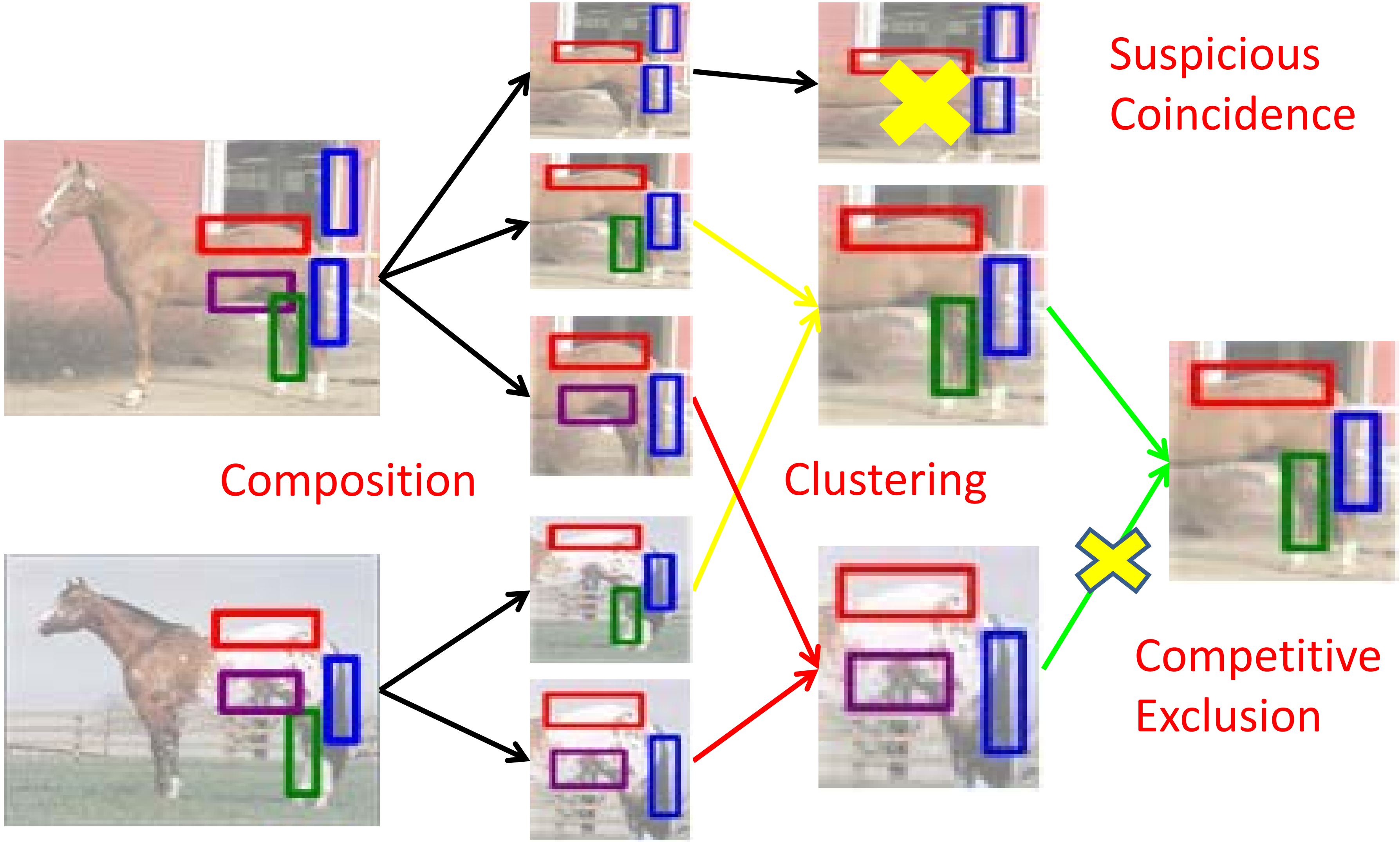
Combinatorial Explosion problem

The Dictionary: From Generic Parts to Object Structures

- Unified representation (RCMs) and learning
- Bridge the gap between the generic features and specific object structures



Bottom-up Learning



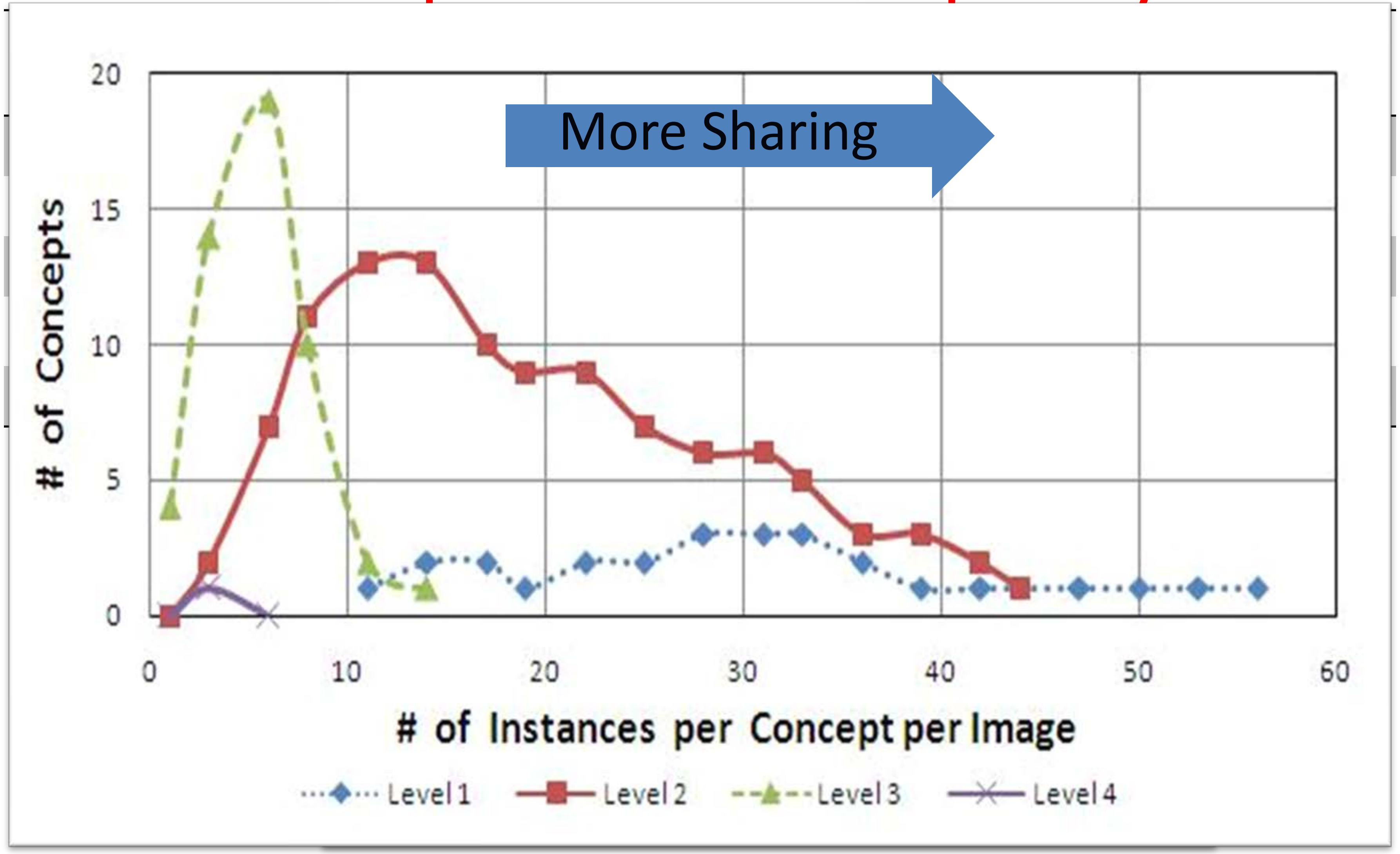
Suspicious Coincidence

Composition

Clustering

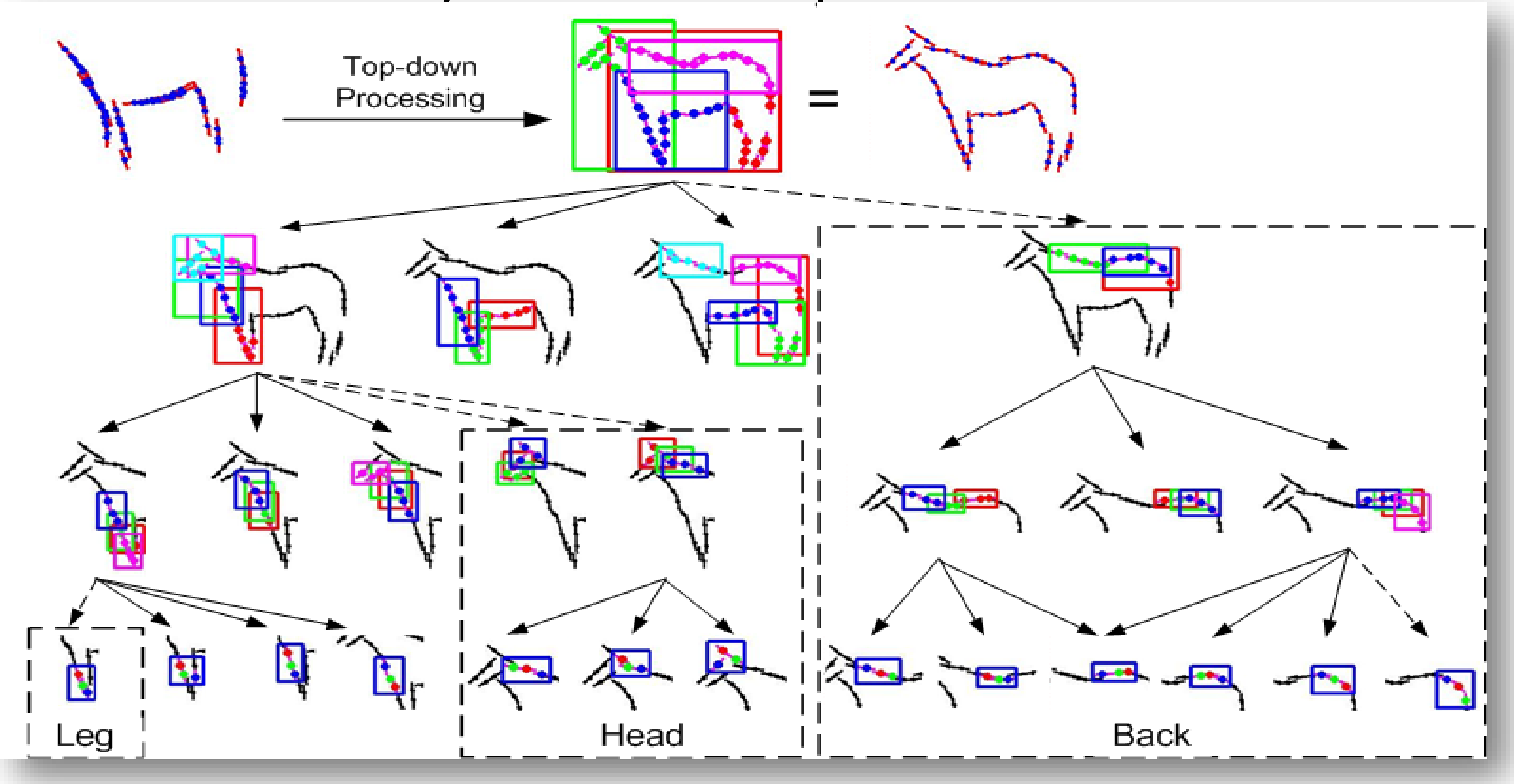
Competitive Exclusion

Dictionary Size, Part Sharing and Computational Complexity



Top-down refinement

- Fill in missing parts
- Examine every node from top to bottom



Part Sharing for multiple objects

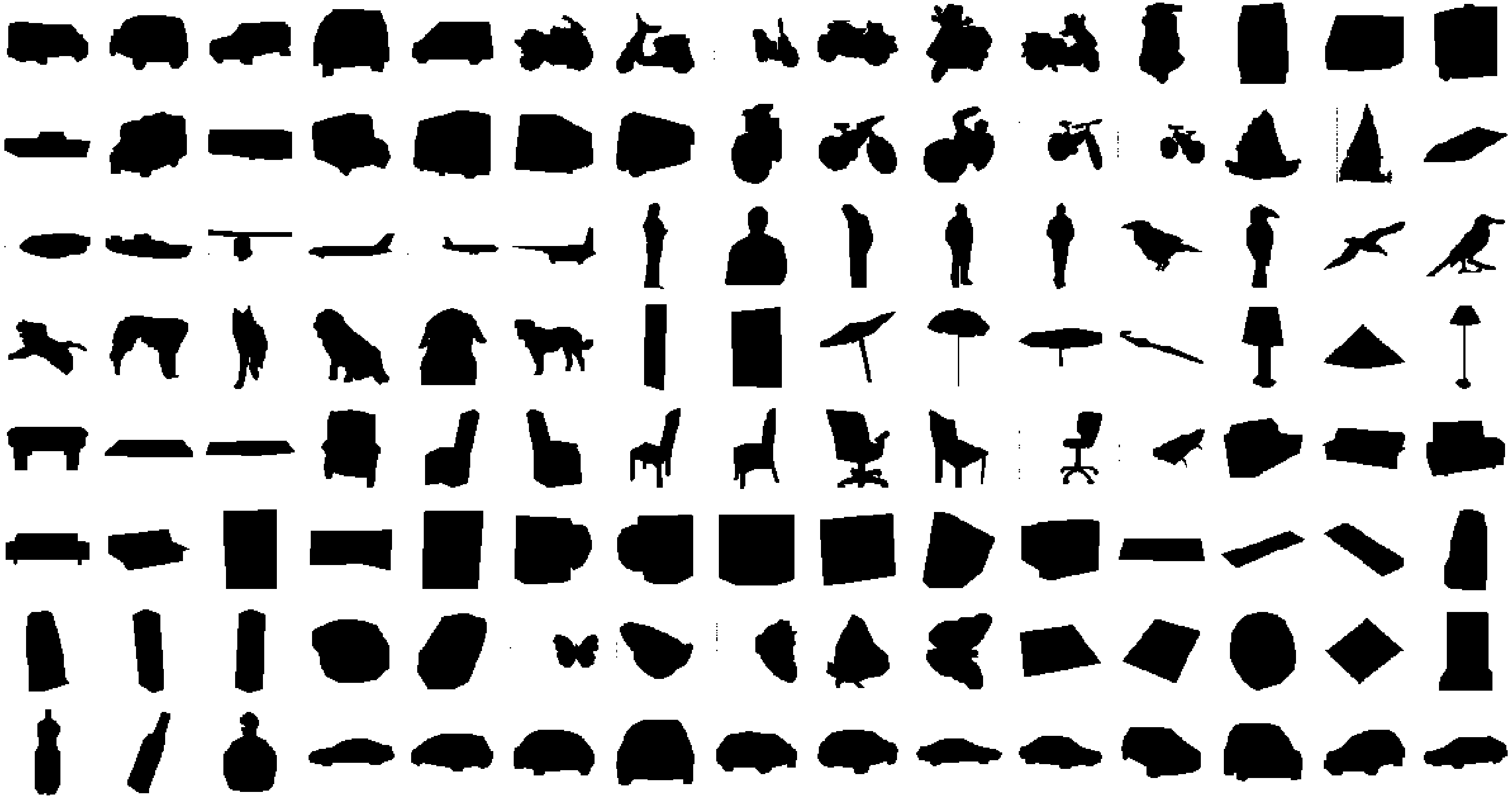
Strategy: share parts between different objects and viewpoints.

Learning Shared Parts

- Unsupervised learning algorithm to learn a hierarchy of parts shared between different objects.
- Structure Induction – learning the graph structures and learning the parameters.
- “I reject this paper because the authors did not say how they choose the number of layers in their hierarchy and how they trained their classifier.” *The whole point of the paper was that there was no classifier to be trained and the algorithm stopped automatically when it found no more suspicious coincidences!*

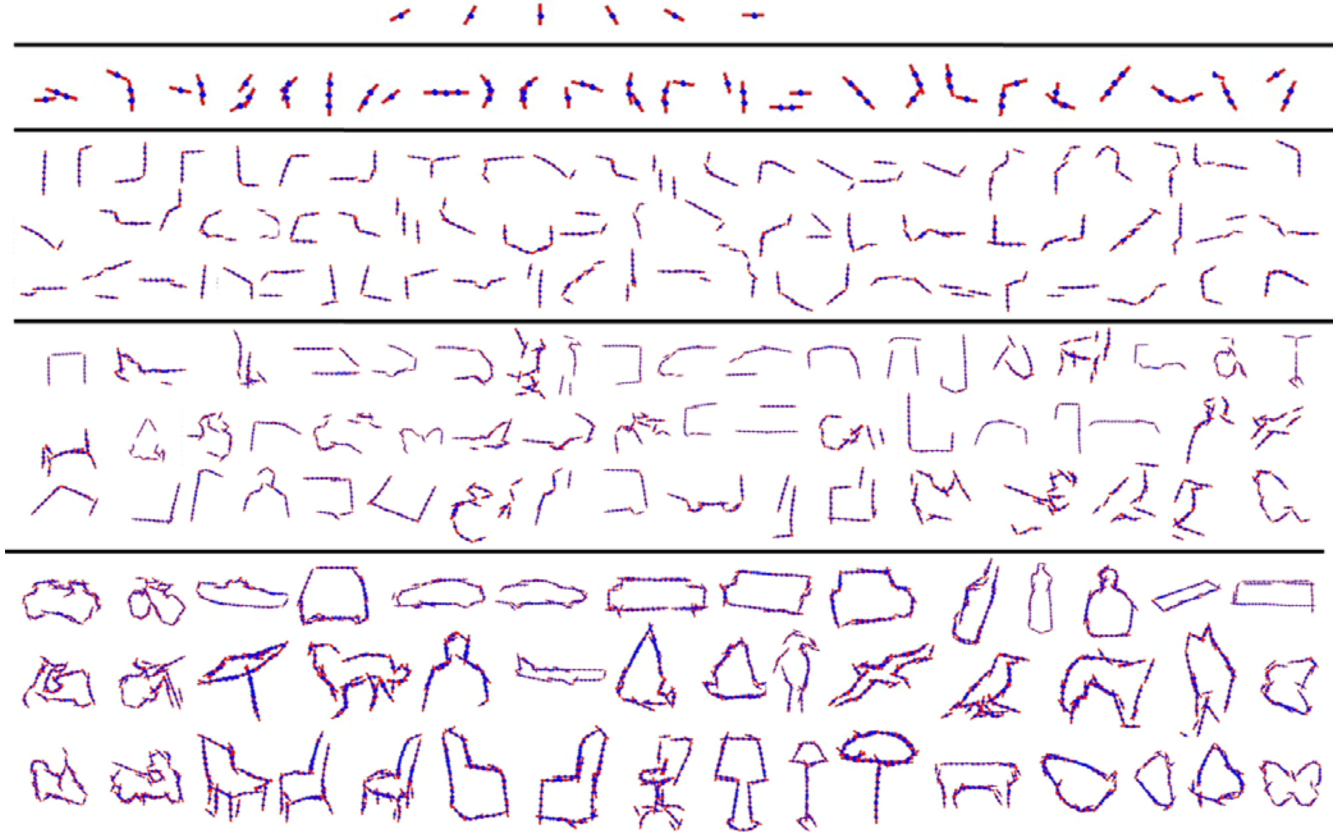
Many Objects/Viewpoints

- 120 templates: 5 viewpoints & 26 classes

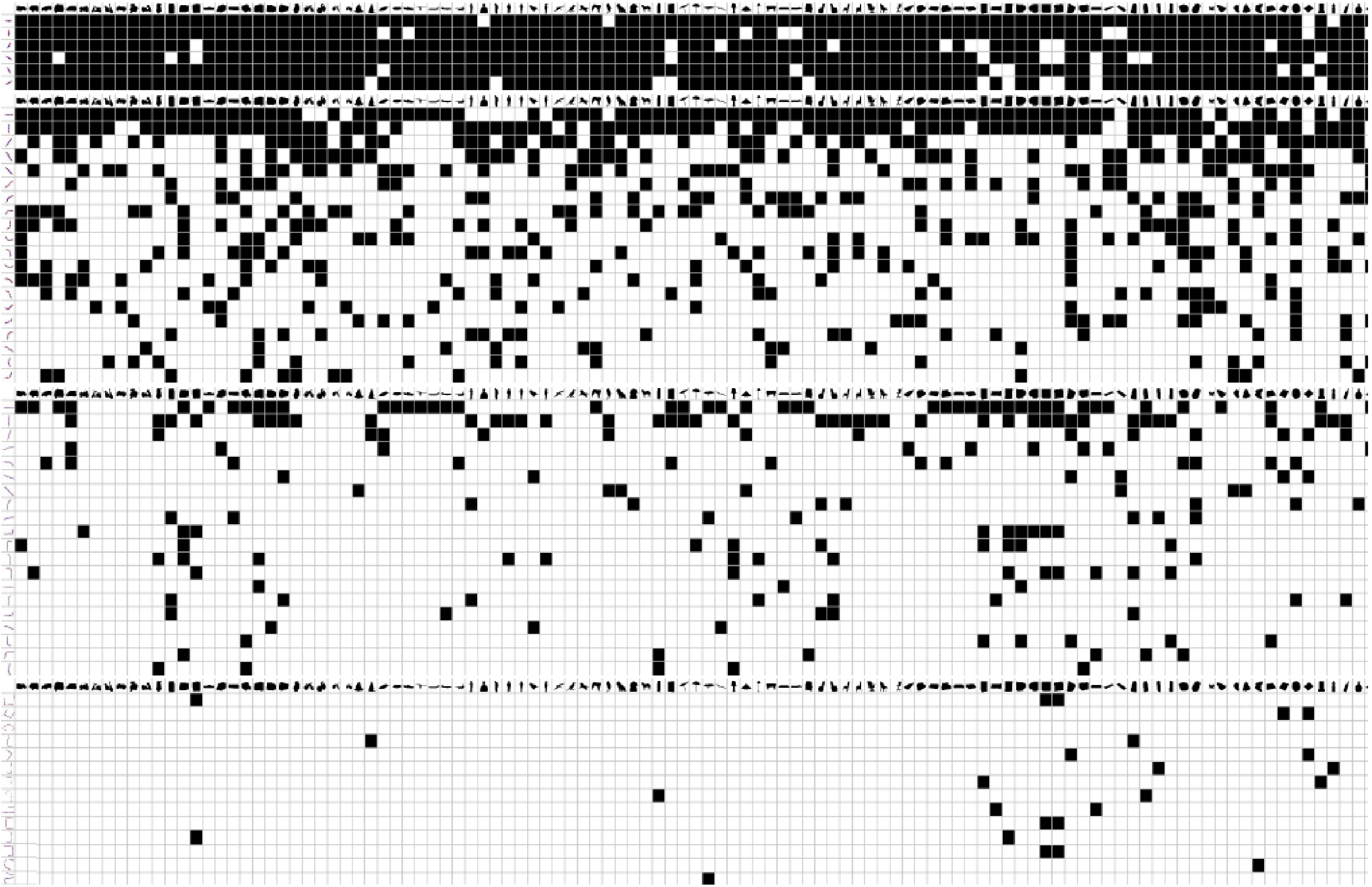


Learn Hierarchical Dictionary.

- Low-level to Mid-level to High-level.
- Learn by suspicious coincidences.

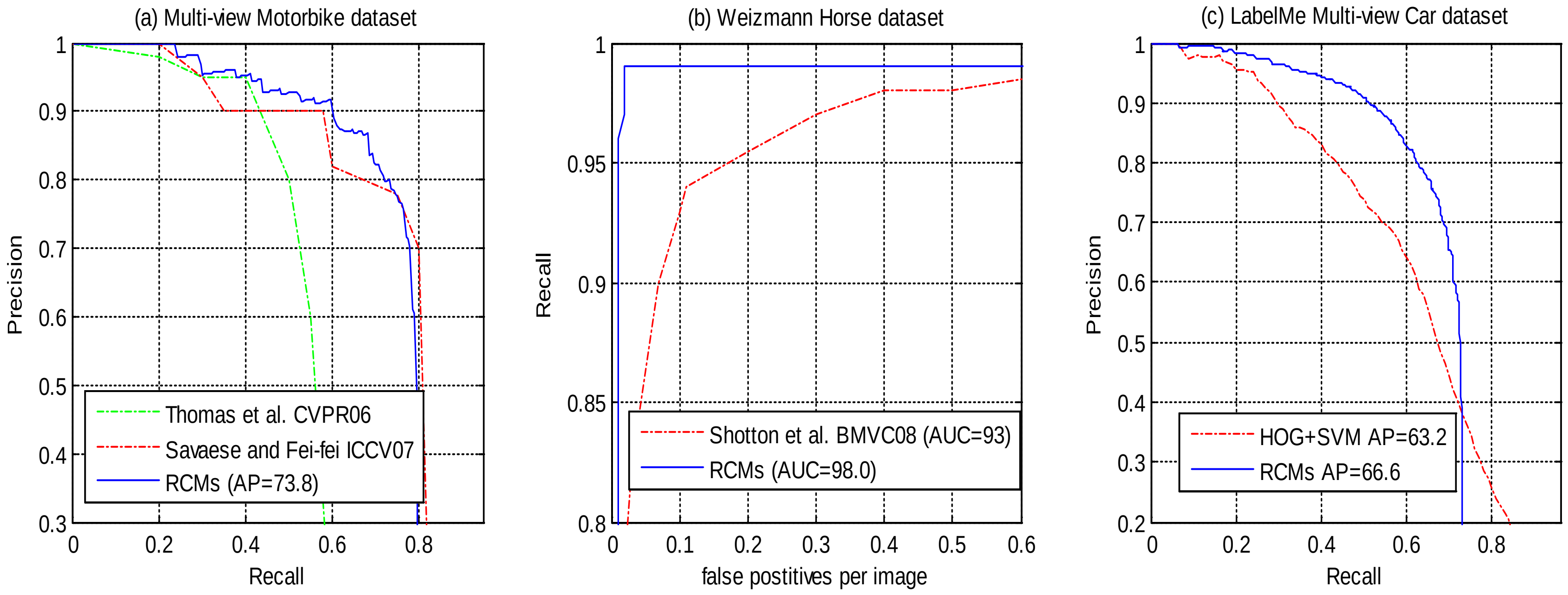


Part Sharing decreases with Levels



Multi-View Single Class Performance

- Comparable to State of the Art (in 2010!).
- Conceptually very attractive, but difficult to implement.



(a) Multi-view Motorbike dataset

(b) Weizmann Horse dataset

(c) LabelMe Multi-view Car dataset

Conclusion

- Learning Hierarchical Models and Architectures is an important goal computer vision. Current vision algorithms lack semantic structure, limits generalization.
- Relying on differentiation limits the class of architectures you can learn.
- There is classic (i.e. pre- Deep Net) work on hierarchical cluster learning to learn hierarchical compositional models with semantic structures.

Brief References

- S. Fidler and A. Leonardis. Towards scalable representations of object categories: Learning a hierarchy of parts. CVPR. 2007.
- S.C. Zhu and D.B. Mumford. A Stochastic Grammar of Images. Now Publishers Inc; 2007.
- Long (Leo) Zhu, Chenxi Lin, Haoda Huang, Yuanhao Chen, Alan Yuille. Unsupervised Structure Learning: Hierarchical Recursive Composition, Suspicious Coincidence and Competitive Exclusion. ECCV 2008.
- Long (Leo) Zhu, Yuanhao Chen, Antonio Torralba, William Freeman, Alan Yuille. Part and Appearance Sharing: Recursive Compositional Models for Multi-View Multi-Object Detection. CVPR 2010.