(1) Compositional Models: Complexity of Representation and Inference

Note Title 11/24/2013

Beyond Feedforward Models:

"something about generative models"

(I) General Comments: Data Driven Models and Representations.
Feedforward and Feedback Architectures.

(II) Unsupervised Learning of Representations and Hierarchies.

(III) Complexity Results on Representations and Inference.

(2) (I) General Comments: Data Driven Models.

Note Title 11/23/2013

In the last 10 years there has been considerable success using data driven models.

- Face Detection. (Viola & Jones)
  - Text Detection (Chen & Yuille)
- Pascal challenge. Object Detection. DPM's. (Felzenszwalb, Ramanan, McAllester) . L. Zhu et al.
- Image Net. DBN's Krizhevsky et al.

Question: Are datasets big and representative enough? Will results scale? Yes and No.

(3) (I) General Comments : Representations

Vision addresses an enormous range of tasks:

Single object: detect in clutter and with occlusion, detect parts and boundaries, estimate 3D structure, reason about pose.

multiple objects: positions relative to each other, occlusion relations, social interactions, etc.

Scene structure: ground plane, Manhattan world structure (if appropriate), positions of objects in scene, background stuff (sky, water), surfaces, geometry, motion.

Visual system: needs to compute rich representations from images.

(4) (I) General Comments: Feedforward and Feedback

- Classic Feedforward Theories:
  Marr: Primal Sketch → 2.5D Sketch → 3D Rep
- Hierarchical Feedforward Models: (Invariances)
  Fukushima, HMax, Deep Belief Networks, M-Theory

Feedforward and Feedback: Generative Models

- Analysis by Synthesis – Mumford & Grenander.
- DDMCMC – Tu & Zhu, Tu, Chen, Yuille, Zhu
  feedforward proposals validated by feedback models.

Ullman: Streams Counter-Streams

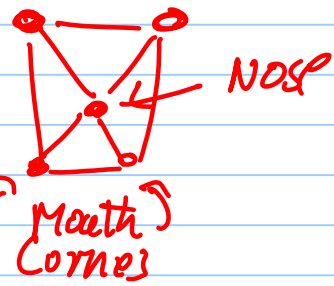
But do generative models require feedback?

(5) (II) Unsupervised Learning of Representations and Hierarchies

Note Title 11/23/2013

Starting Point : Pictorial Structures. [Fischler & Erschlager 1973]

Model object (Face) in terms of parts



$G = (V, E)$ Graph

nodes + edges

State variable $\{\underline{x}_\mu : \mu \in V\}$

energy $E[\{\underline{x}_\mu\}] = \sum_{\mu \in V} \underline{\lambda}^\mu \cdot \underline{\phi}(\underline{x}_\mu) + \sum_{(\mu,\nu) \in E} \underline{\lambda}^{\mu\nu} \cdot \underline{\psi}(\underline{x}_\mu, \underline{x}_\nu)$

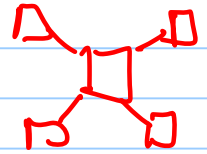
data term. ; spatial relations

Deformable Template :

(G) (II) Unsupervised Learning of Representations and Hierarchy.

These models are very successful — often by putting many features into the data term.

Examples.
- von der Malsburg — Face Detection "Neural". Wiskott., Neven
- Coughlan, Yuille — Hands (Dynamic Programming)
- Felzenszwalb, Ramanan, McAllester — Star Model, DPM (Pascal Detection)

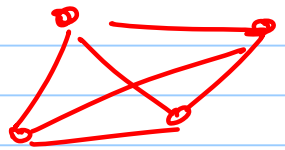


(7) (II) Unsupervised Learning of Representations and Hierarchies

Unsupervised Learning

Constellation Models. Caltech.

- Weber, Perona
- Fergus et al.



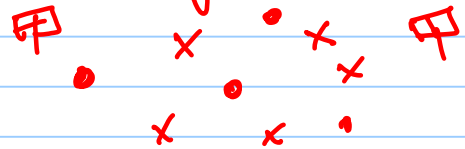
Fully connected ?

Flat Models → no composition

(8) (II) Unsupervised Learning of Representation and Hierarchies

1. Zhu et al: Learn Mixtures of Models
Data with clutter

- Take Image → Extract and Represent Interest Points (IPs)



· Interest Point types. (~200).

- Task - Learn a Generative Model.

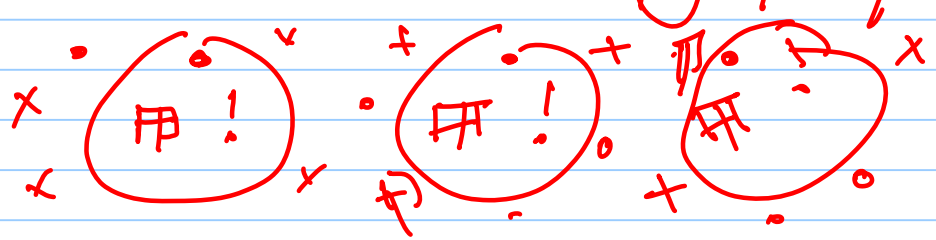
Don't know: (i) how many objects
(ii) how many IP's in an object
(iii) which IP's are object or background.

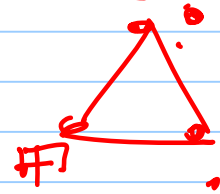
(9) (II) Unsupervised Learning of Representations and Hierarchies

Strategy: Greedy search over space of models

Initialize: Default model all IP's generated independently (1.i.D)

Cluster: Identify frequent triplets





Gaussian Distribution over spatial positions

"Suspicious coincides". Barlow

"meaningful alignments" Ullman.

Note: Gaussian over internal angles invariant to rotation/scale.

[16] (II) Unsupervised Learning of Representation and Hierarchies

Grow Model by Adding Triplets:

Current Model :

Image. N background point B

+ 3M points generated by △ Triplet Model.

IP's generated by i.i.d. background or triplet.

B △

↓ ↓ Rep by B + △

Grow to B + △ or B + △ + ▽

↑ more complex object

× new object

(11) Unsupervised Learning of Representations and Hierarchies

Better Encoding of Data:

- Cost of Encoding Data by $\sum \log P(\cdot)$ only Default!
- Cost of Encoding Data by $B + \Delta$ $\sum \log P \dots + T \times$ (overhead)

Model Selection: Stop when adding "new" object or growing bi

(12) (II) Unsupervised Learning of Representations and Objects

Learn Representations → Unsupervised.
→ I.P.'s only (less interesting)

Learning in Cocktail Party.

Unknown no. of speakers + Background Noise.

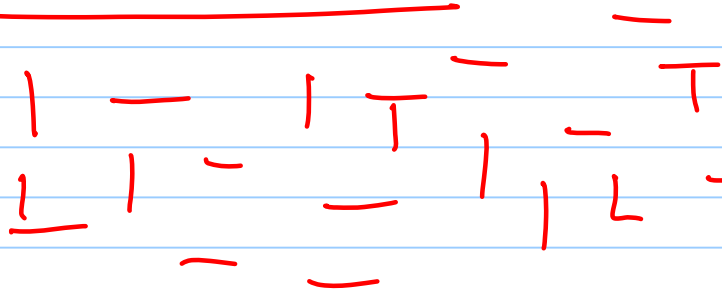
Can do non-trivial visual tasks

E.G. Cars from multiple viewpoints: Motlaghi & Yuille

Search over Space of Graphical Models: Kemp & Tenenbaum (IVAE)

(13) (II) Unsupervised Learning of Representation & Hierarchies.

L. Zhu et al. 2008, 2010 — Hierarchies of Edges



Basic Dictionary

vertical / horizontal bar.

First Stage: same as before

cluster: identify compositon T and L.

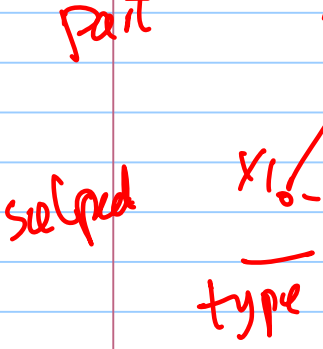
don't do model selection → don't force decision on T

require T + B, L + B to encode better than B

(14) (II): Unsupervised Learning of Representations and Hierarchies

Composition: Part-Subpart

S. Geman
Mumford & Desolneux



$$P(x_1, x_2 | x) = \delta(x - \tfrac{1}{2}(x_1 + x_2))\, h(x_1, x_2 : \lambda)$$

$$h(x_1, x_2 : \lambda) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_1 - x_2 - \mu)^2 / 2\sigma^2}$$

$$\lambda = (\mu, \sigma)$$

T = "|" + "—" + "spatial relation."
$\underline{\lambda}_1 = (\mu_1, \sigma)$

L = "|" + "—" + "spatial relation"
$\underline{\lambda}_2 = (\mu_2, \sigma)$.

"Visual Concepts"
Deformable Models

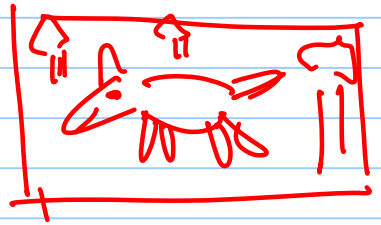
(15) (II) : Unsupervised learning of Repetition and Hierarchies

Second Stage:

- Dictionary of low-level features. "—" "|" / , \
- Visual Concepts level 1. — \ T L
- Apply same clustering procedure to visual concepts at level 1 (at this stage resolve ambiguities in T)
- Obtain Visual Concepts level 2.

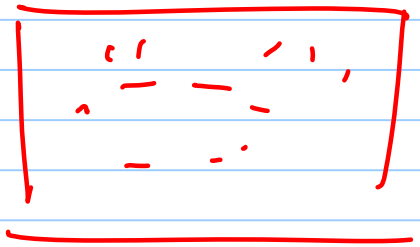
Repeat → until you stop finding suspicious coincidences

(15½) How to learn the representation? (ECCV'08)



→ Edge Detection.

Cocktail Party



Strategy: Low-Level Dictionary

Look for compositions of triples that happen frequently with spatial variability (Gaussian + mean).

- Visual Concepts
- Look for Compositions of Visual Concepts.

Matter of Encoding.

Justification → Parallel Search over encodings of the image.

(15 3/4) Keep Grouping until you cannot find any component.

Visual Concepts 5

Visual Concepts 4

Visual Concepts 3

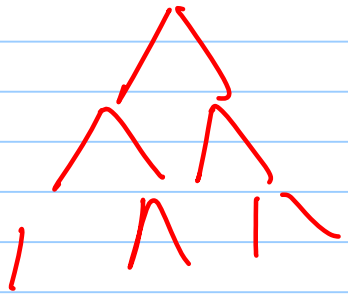
~~Visual Concepts~~ 2

Dictionary 1

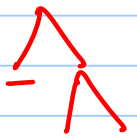
Weigmin //

Model. Selection.

Start. with Top Visual Concept

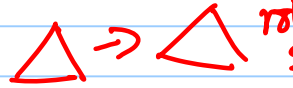


Add Extra

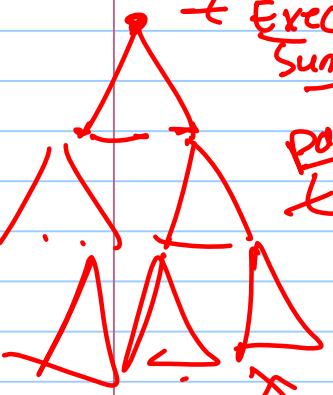


better encoding Scno.

After learning → relax 2/3 rule

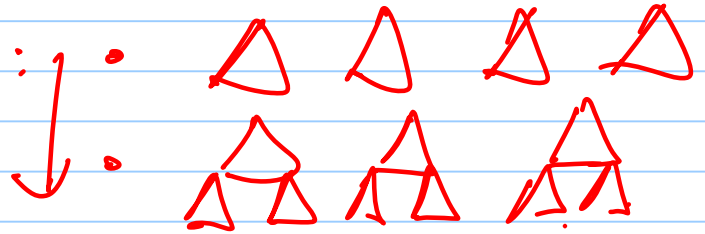


16) (II) Unsupervised Learning Representation and Hierarchies



- Learn hierarchical graphical models
- Hierarchical Distributed Representation of Object.
- Coarse - Executive Level Summary
  Parent state is invariant to details of children.
- Finer scale details at lower nodes
  positions of parts, boundaries

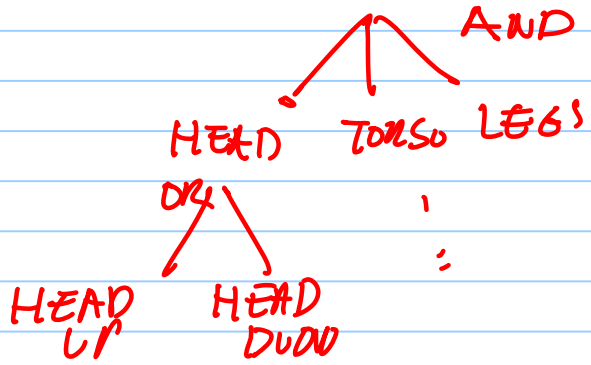
Hierarchy of Visual Concepts:



(17) (II) Unsupervised Learning Representations and Hierarchies

Note: These types of models already existed as hand specified models.

AND-OR Graphs Baseball Players

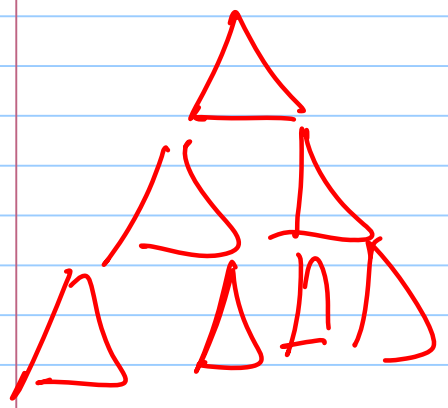


L. Zhu et al.
S. Zhu et al.
S. Geman.
C. Williams.

(48) (II) Unsupervised Learning of Representations and Hierarchies

Exact Inference on Generative Models.

Feedforward gives executive summary
Feedback resolves low level ambiguities



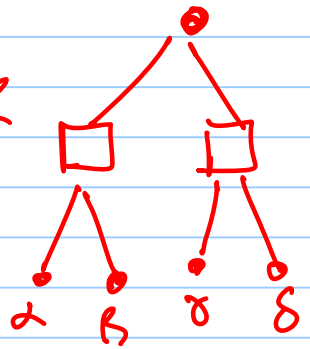
"Inference on Generative Models can be done very fast."

Feedforward propagates up low-level hypothesis / ambiguous
At high-levels that is sufficient context to disambiguate.
Top-Down uses high-level context to resolve low-level ambiguities
(Binding / Linking)

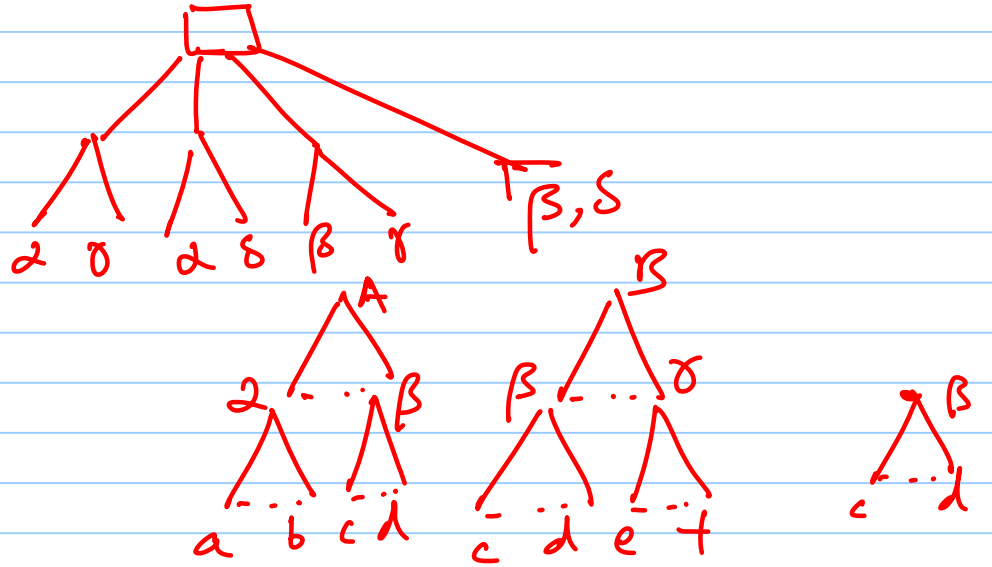
(14) (II) Unsupervised Learning Representations and Hierarchies

Convert.

AND-OR graph



→ OR of AND Graphs



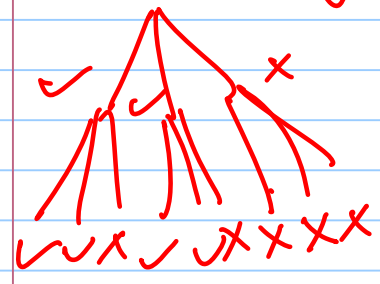
Why Hierarchies ?

Part-Sharing.

(20) (II) Complexity Results: Representation and Inference

Part Sharing : Efficiency of Representation and Inference (and Learning)

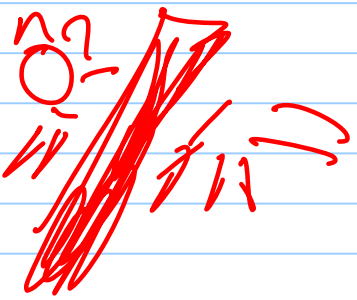
Also robustness to missing parts.



2/3 rule.

only need to detect 4 out of 9 sub-sub-parts

many parts missing at fine level.

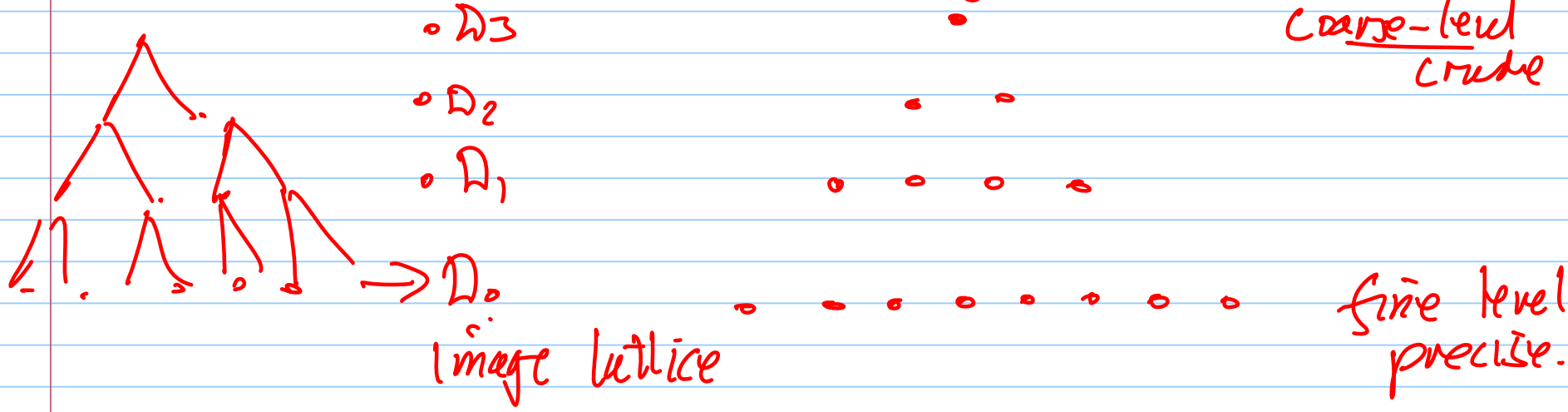


few points missing at high level

(21) (III) Complexity Results: Representation and Inference

Inference: Computation is done by Dynamic Programming
Can compute exact no. of computations.

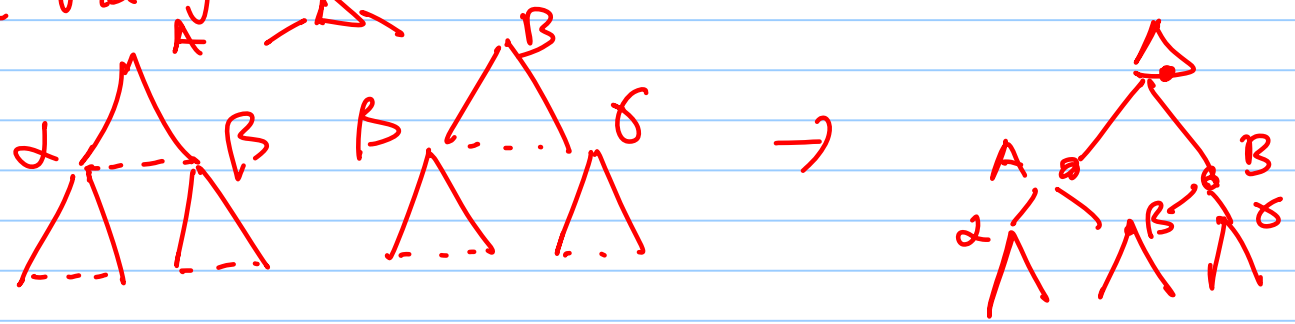
First: "Quantize" the states of object model.



(2 (½)

Note: Inference is Exact. with Sharing

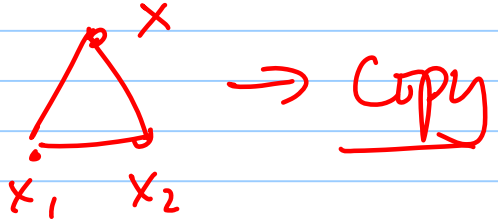
Despite Many Closed Loops.



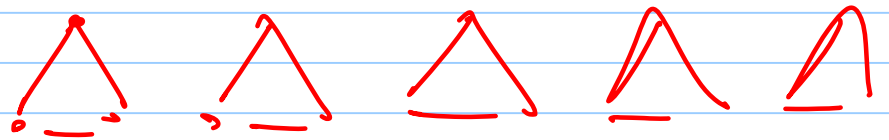
2. Valiant Circuits of the Mind.

Fundamental Problem → Can we devise the structure of the visual cortex from first principles.

(22) (III) Complexity of Representation and Inference

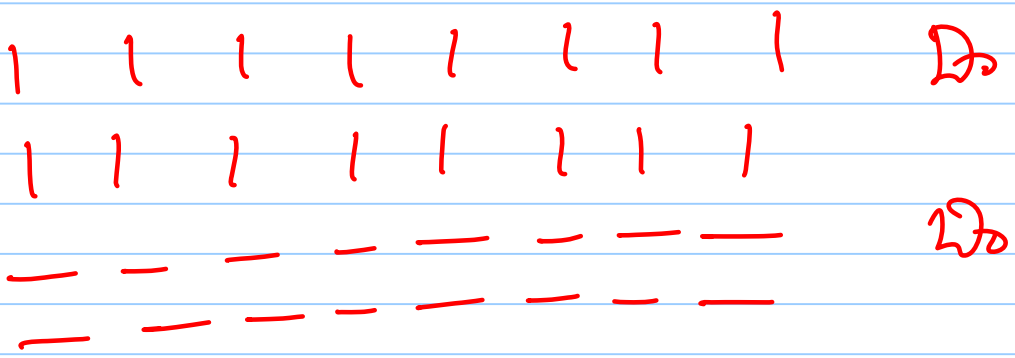


→ Copy



Lowest-level ⊢

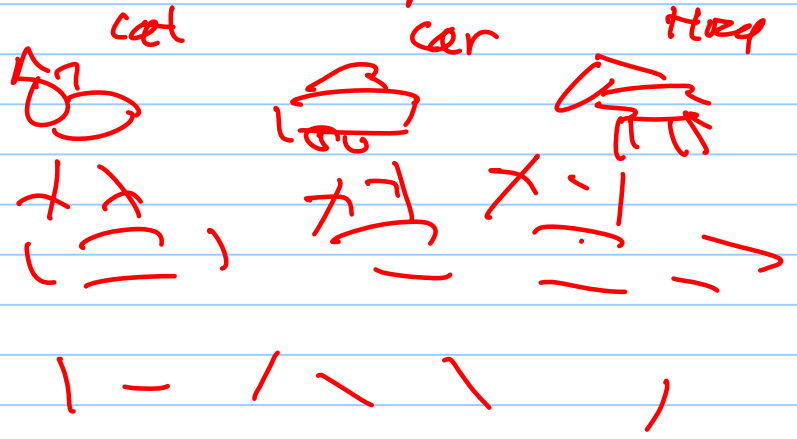
"Columns"



Encode higher-level concepts sparsely in position

(23) (III) Complexity of Representation and Inference

Toy Model of Cortex: Non-linear receptive fields



Sparse Activity -

(24) Complexity: Representation and Inference

Fundamental Problem of Vision → Complexity

How to learn, represent, access all possible objects.

rapidity?

Compositional Hypothesis → possible because objects are compound in term of visual concepts (parts) constructed from more elementary visual concepts (sub-parts).

Hierarchical Representation → Executive Summary → Shared Parts.

Visual Architecture → Neural Model.

L. Voliant.

· Circuits of the Mind.

(25) Summary.

- Complexity of Vision: Fundamental Problem?
  (i) Tasks
  (ii) Images.
- Composited Models → Explicit Representation of Parts/subparts including spatial relations; Part sharing offers a way to address complexity.
  Extensible ⇒ Object Appearance, Scenes, 3D, ...