Note Title 11/24/2013

(1) Compositional Models: Complexity of Representation and Inference

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. DPM's. (Felzenszwalb, Ramanan, McAllester)
  Object Detection . 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 approp), positions of objects in scene, background stuff (sky, water), surfaces, geometry, motion.

Visual system: needs to compute rich representations from images.

(4) (I) <u>General Comments</u> : <u>Feedforward and Feedback</u>

- 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.

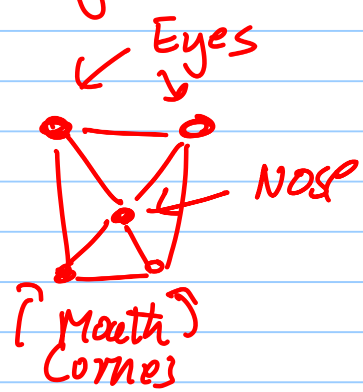
$\boxed{\text{Ullman: Streams Counter-Streams}}$

<u>But</u> do generative models require feedback?

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

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

Model Object (Face) in terms of parts



$G = (\mathcal{V}, \mathcal{E})$ Graph

nodes, + edges

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

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

data term. spatial relations

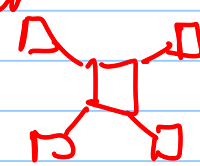
Deformable Template :

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

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

Examples:

- Von der Malsburg – Face Detection "Neural".
  Wiscott., Neven
- Coughlan, Yuille – Hands (Dynamic Programming)
- Felzenszwalb, Ramanan, McAllester – Star Model, DPM
  (Pascal Detection)

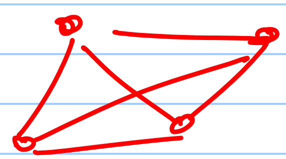


(7) (II) Unsupervised Learning of Representations and Hierarchies

Unsupervised Learning

Constellation Models. Caltech.

- Weber, Perona
- Fergus et al.



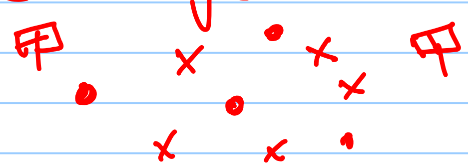
Fully connected ?

Flat Models → no composition

(8) (II) <u>Unsupervised Learning of Representation and Hierarchies</u>

<u>1. Zhu et al:</u> Learn Mixture of Models
Data with clutter

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



· Interest Point types. ⊞ • ×
(~200).

- Task – Learn a Generative Model.

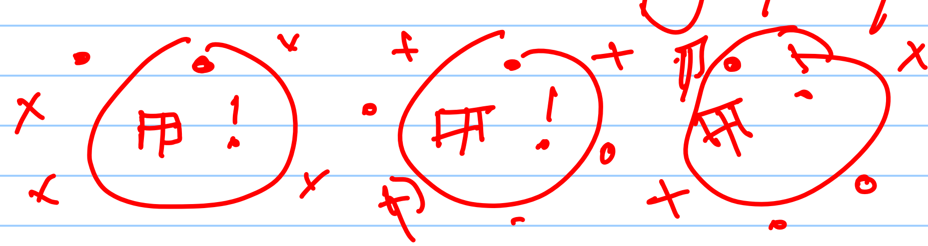
Don't know : (i) how many objects
(ii) how many IP's in an object
(iii) which IP's are object or background.

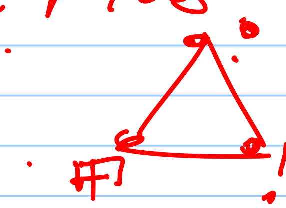
(9) (II) <u>Unsupervised Learning of Representations and Hierarchies</u>

<u>Strategy</u>: Greedy search over space of models

<u>Initialize</u>: Default model all IP's generated independently (I.I.D)

Cluster: Identify frequent triplets





Gaussian Distrib. over spatial position

"Suspicious coincides". Barlow

"meaningful allignments" Morel.

'...." Ullman.

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

[18] (II) <u>Unsupervised Learning of Representation and Hierarchies</u>

<u>Grow Model by Adding Triplets:</u>

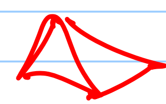
<u>Current Model</u>:

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

<u>Image</u>: N background point B
+ 3M points generated by $\triangle$ Triplet Model.

$\cdot B$ $\quad \triangle$

$\downarrow \quad \downarrow$ <u>Rep</u> by $B + \triangle$

<u>Grow</u> to $B +$  or $B + \triangle + \nabla$ $\times$ new object

more complex object

(11) <u>Unsupervised Learning of Representation and Hierarchies</u>

Better Encoding of Data:

- Cost of Encoding Data by B only Default

  $\sum_i \log P(\cdot)$

- Cost of Encoding Data by $B + \Delta$

  $\sum \log P \ldots + T_{\text{model}}$

<u>Model Selection:</u>

Stop when adding "new" object or growing bi

(12) [II] <u>Unsupervised Learning of Representations and Objects</u>

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

<u>Learning</u> in <u>Cocktail Party</u>.

Unknown no. of speakers + Background Noise.

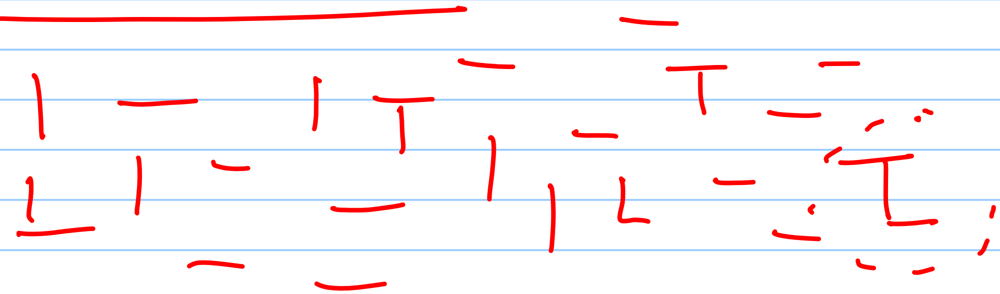
<u>Can</u> do non-trivial vision tasks

<u>E.G.</u> Cars from multiple viewpoints : Mottaghi & Yuille

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

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

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



Basic Dictionary

vertical / horizontal bar.

First Stage: same as before

cluster: identify compositions T and L.

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

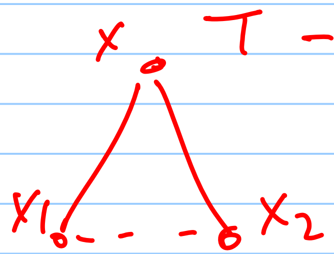
require T + B, L + B to encode better than B

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

Composition: Part - Subpart

S. Geman
Mumford & Desolneux

part



T - type

subpart

— type | type

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

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

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

T = "|" + "—" + "spatial relation"
$\lambda_1 = (\mu_1, \sigma)$

L = "|" + "—" + "spatial relation"
$\lambda_2 = (\mu_2, \sigma)$.

"Visual Concepts"
Deformable Models

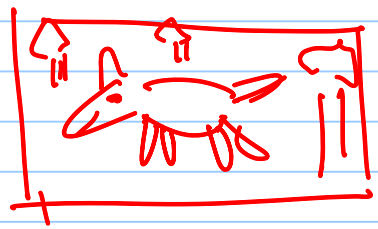
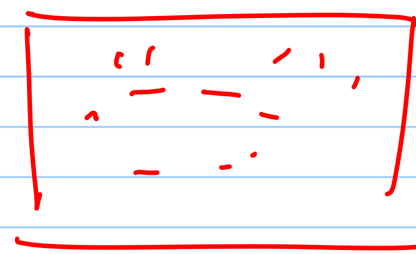
(15) (II) : <u>Unsupervised learning of Representation and Hierarchies</u>

<u>Second Stage:</u>

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

   <u>Repeat</u> → until you stop finding suspicious coincidences

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

 → Edge Detection. 

Cocktail Party

Strategy : Low-Level Dictionary | — \ /

Look for conjunctions 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

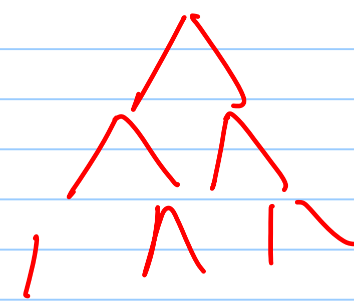
~~Visual Concept~~ 2

Dictionary 1

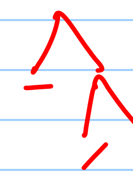
Weizmin //

Model. Selection.

Start. with Top Visual Concept

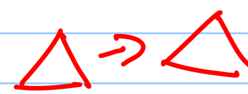


Add Extra



better encoding score.

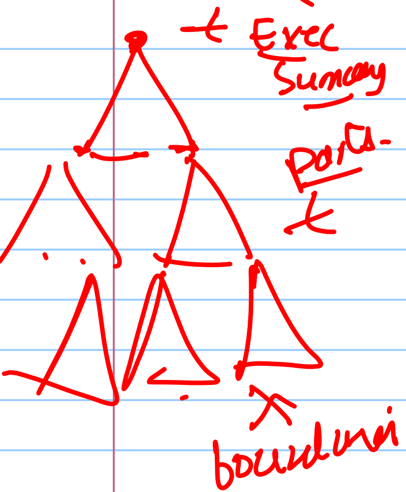
After learning → relax 2/3 rule



rotation scale.

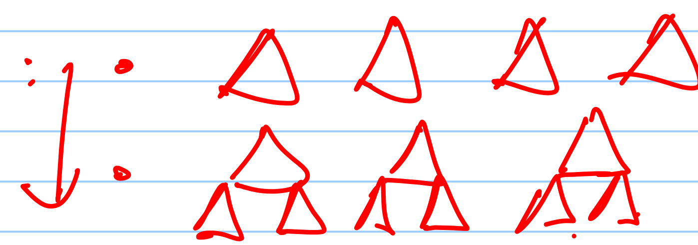
/ 16)

(II) <u>Unsupervised Learning Representation and Hierarchies</u>



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

<u>Hierarchy of Visual Concepts</u> :

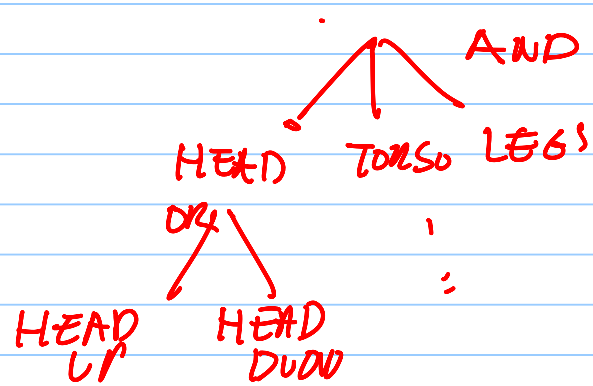


(17) (II) Unsupervised Learning Representations and Hierarchies

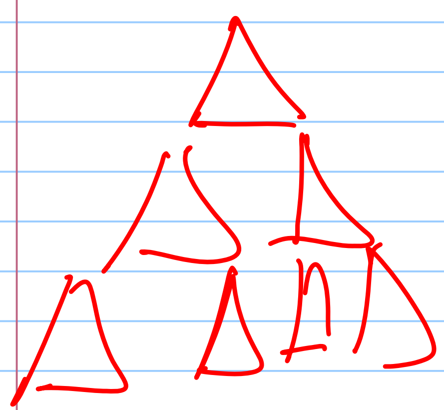
Note: these types of models already existed as hand specified models.

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

AND-OR Graphs Baseball Players



(+8) (II) <u>Unsupervised Learning of Representations and Hierarchies</u>



<u>Exact Inference</u> on Generative Model.

Feedforward gives executive summary.

Feedback resolves low level ambiguities

"<u>Inference</u> on Generative Models can be done very fast."

Feedforward propagates up low-level hypothesis / ambiguous

At high-levels there is sufficient context to disambiguate.

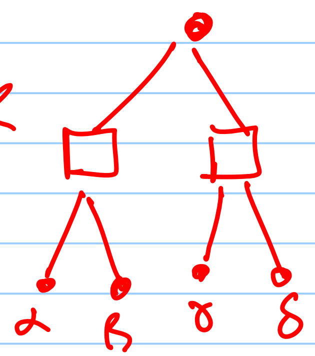
Top-Down uses high-level context to resolve low-level ambiguities

(Binding / Linking)

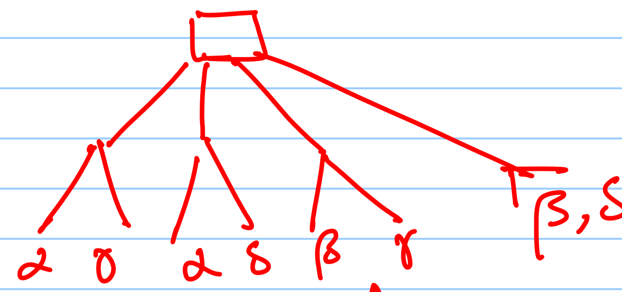
(14) (II) Unsupervised Learning Representations and Hierarchies

Convert.

AND-OR graph

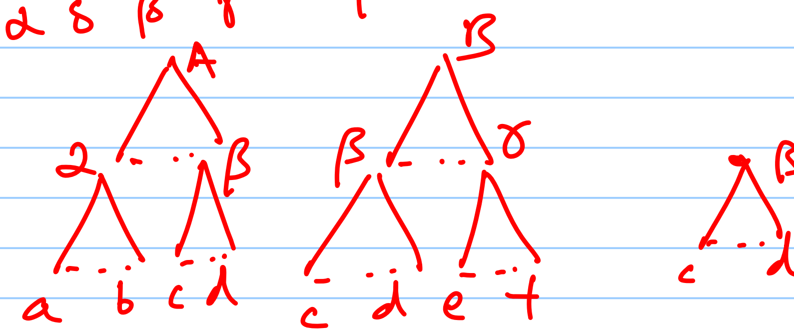


$\Longrightarrow$ OR of AND graphs



Why Hierarchies?

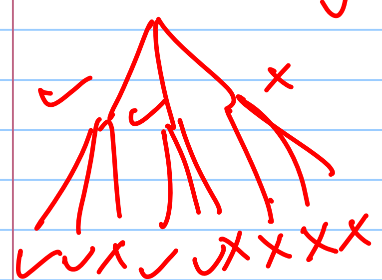
Part-Sharing.



(20) [III] <u>Complexity Results:</u> <u>Representation and Inference</u>

<u>Part Sharing</u> : Efficiency of Representation and Inference (and Learning).

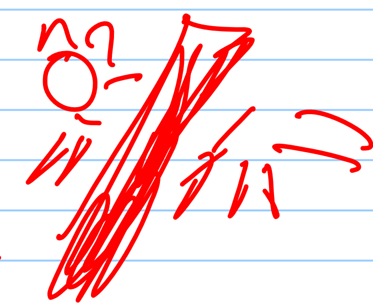
Also robustness to missing parts.



2/<u>3</u> rule.

<u>only</u> need to detect 4 out of 9 sub-sub-parts

many parts missing at fine level.

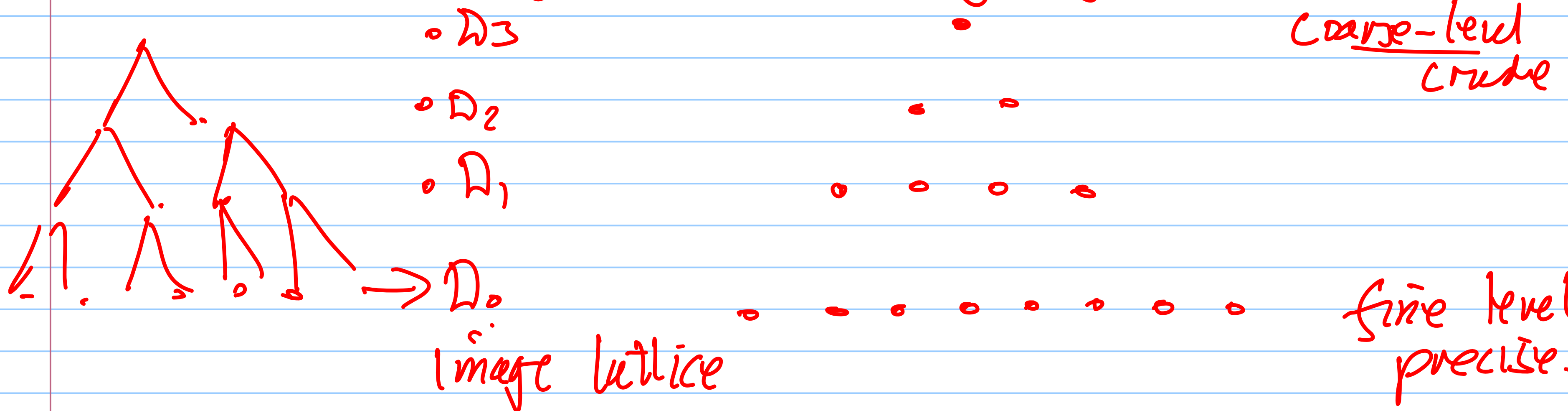


few parts missing at high level

(21) $\boxed{\text{III}}$ <u>Complexity</u> Results: <u>Representation and Inference</u>

<u>Inference:</u> Computation is done by Dynamic Programming
Can compute exact no. of computations.

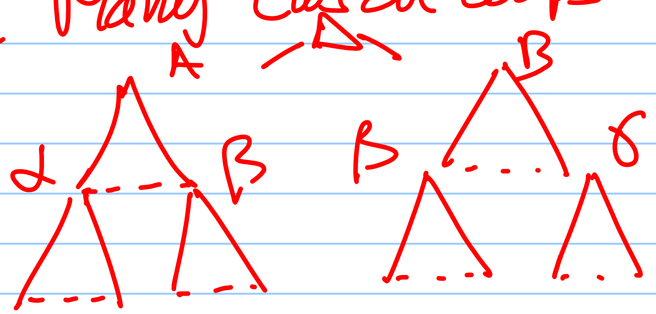
<u>First:</u> "Quantize" the states of object model.



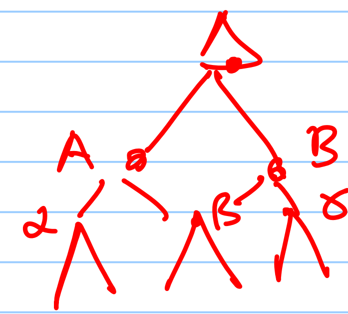
(21½)

Note: Inference is Exact with Sharing

Despite Many Closed Loops.



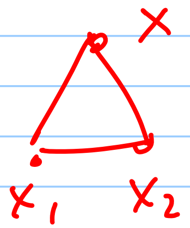
$\rightarrow$



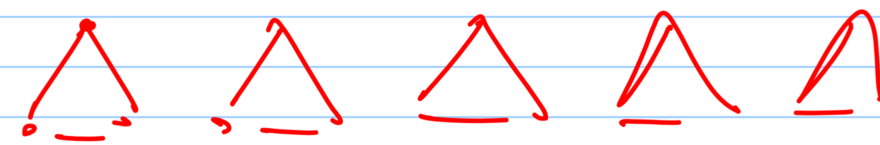
2. Valiant Circuits of the Mind.

Fundamental Problem $\rightarrow$ Can we derive the structure of the visual cortex from first principles.

(22) (III) Complexity of Representation and Inference

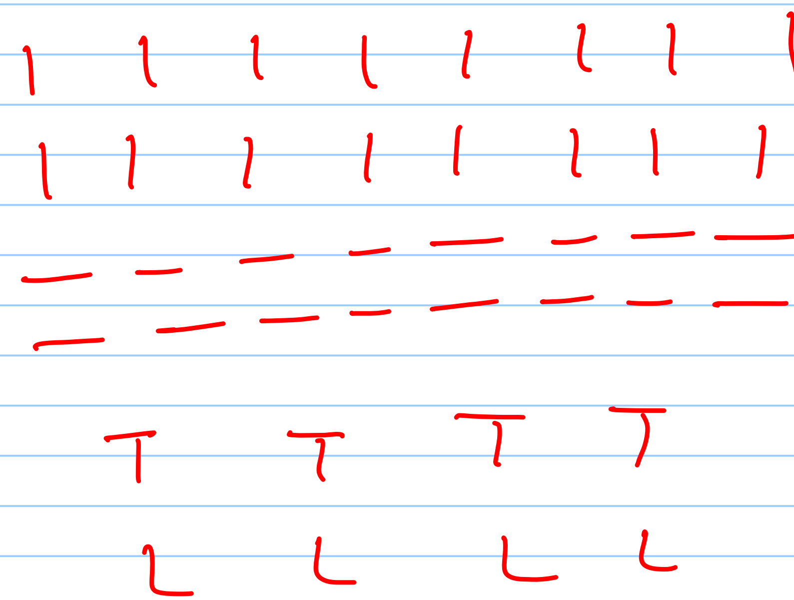


$\rightarrow$ Copy



Lowest-Level 1 —

"Columns"



D_0

D_0

D_1

Encode higher-level concepts sparsely in position

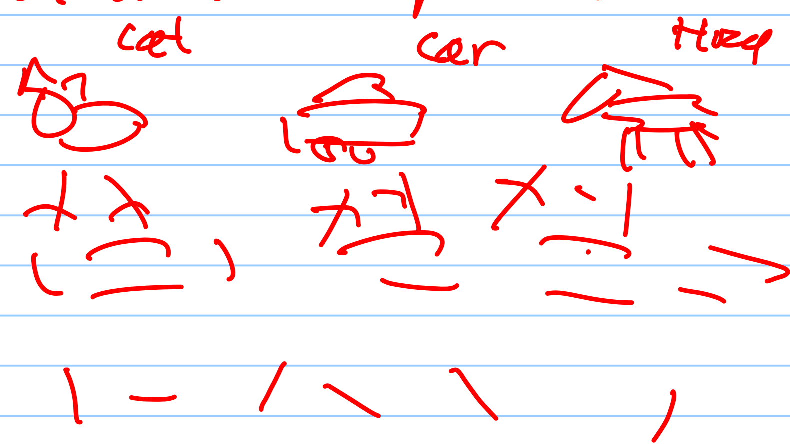
(23) (III) <u>Complexity of Representation and Inference</u>

<u>Toy Model of Cortex:</u> Non-linear receptive fields

<u>Objects</u>

<u>Parts</u>

<u>Low level</u>



<u>Sparse Activity</u>.

(24) <u>Complexity</u>: <u>Representation and Inference</u>.

<u>Fundamental Problem of Vision</u> → Complexity

How to learn, represent, access all possible objects rapidly?

Compositional hypothesis → possible because objects are composed in terms of visual concepts (parts) constructed from more elementary visual concepts (sub-parts).

Hierarchical Representation → Executive Summary
→ Shared Parts.

<u>Visual Architecture</u> → Neural Model.

L. Valiant. Circuits of the Mind.

(25) **Summary.**

- Complexity of Vision: Fundamental Problem?

(i) Tasks

(ii) Images.

- Compositional Models $\Rightarrow$ Explicit Representation of Parts/subparts including spatial relations, Part Sharing

Offers a way to address complexity.

Extensible $\Rightarrow$ Object Appearance, Scenes, 3D, . . .