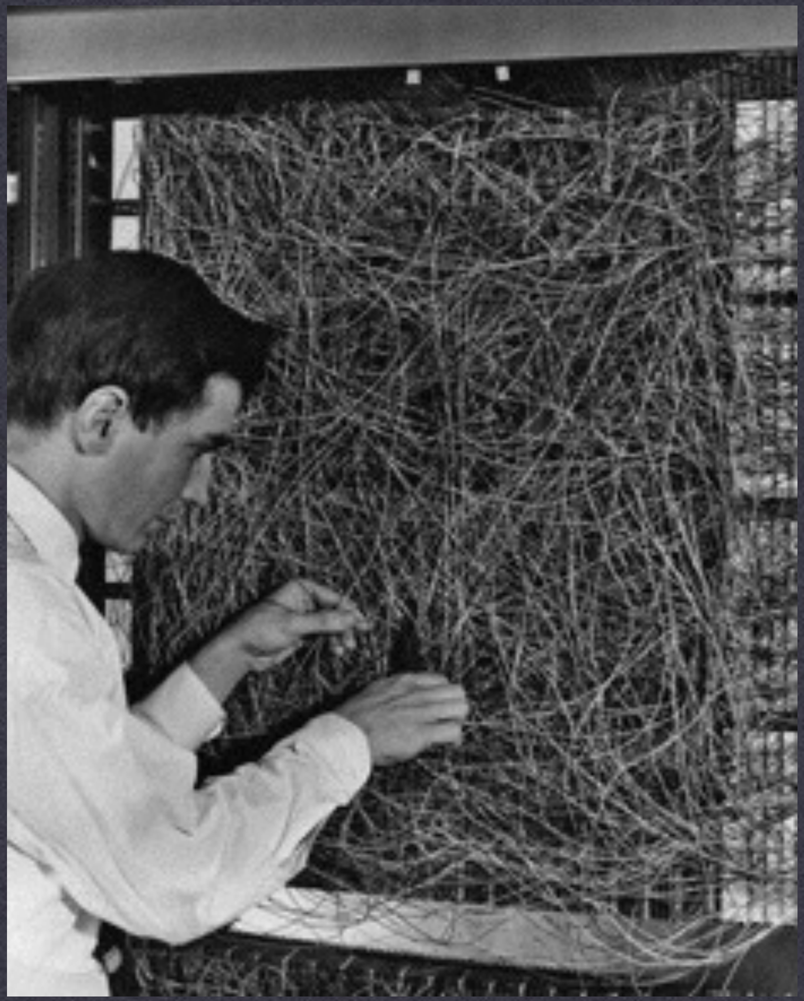
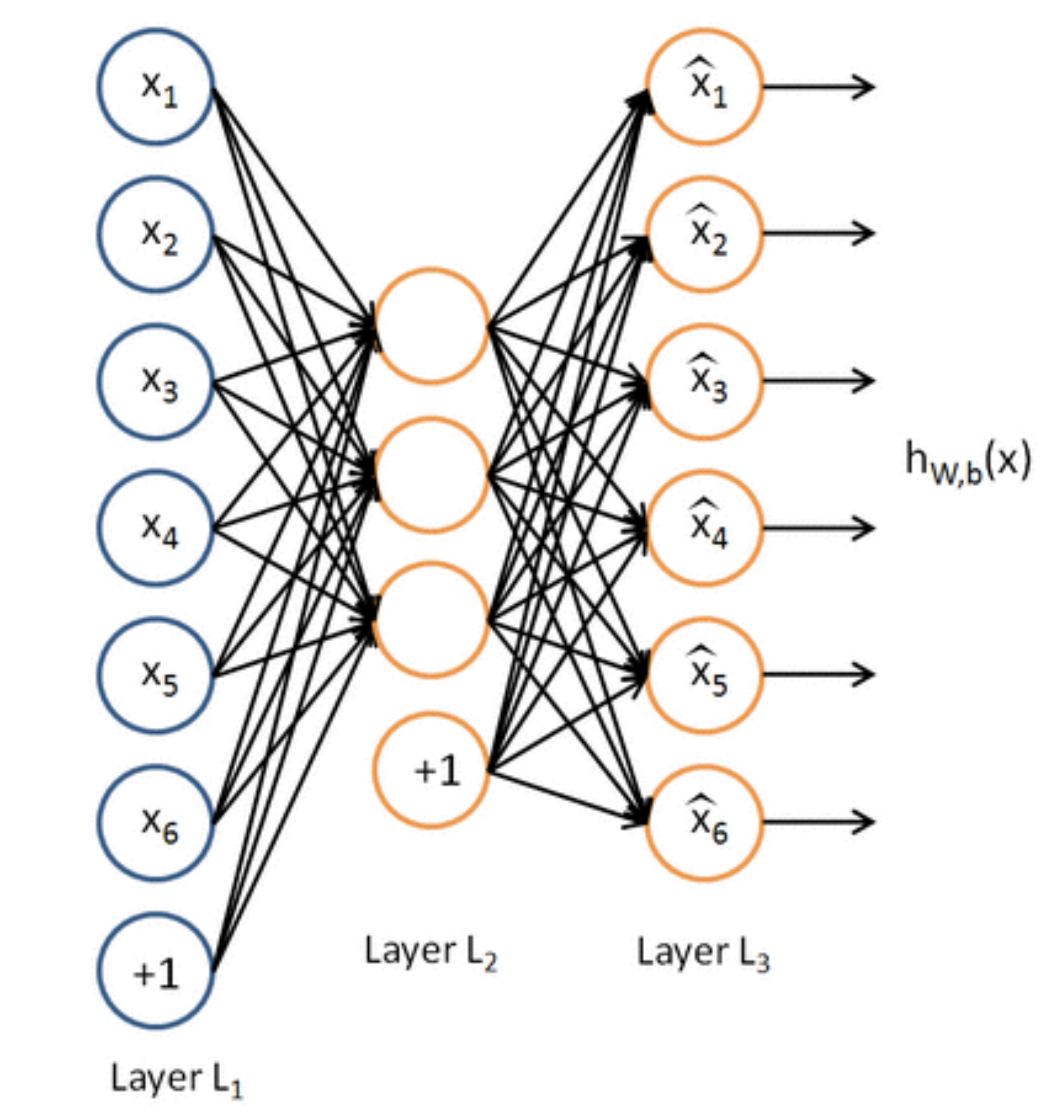


RESTRICTED BOLTZMANN MACHINES

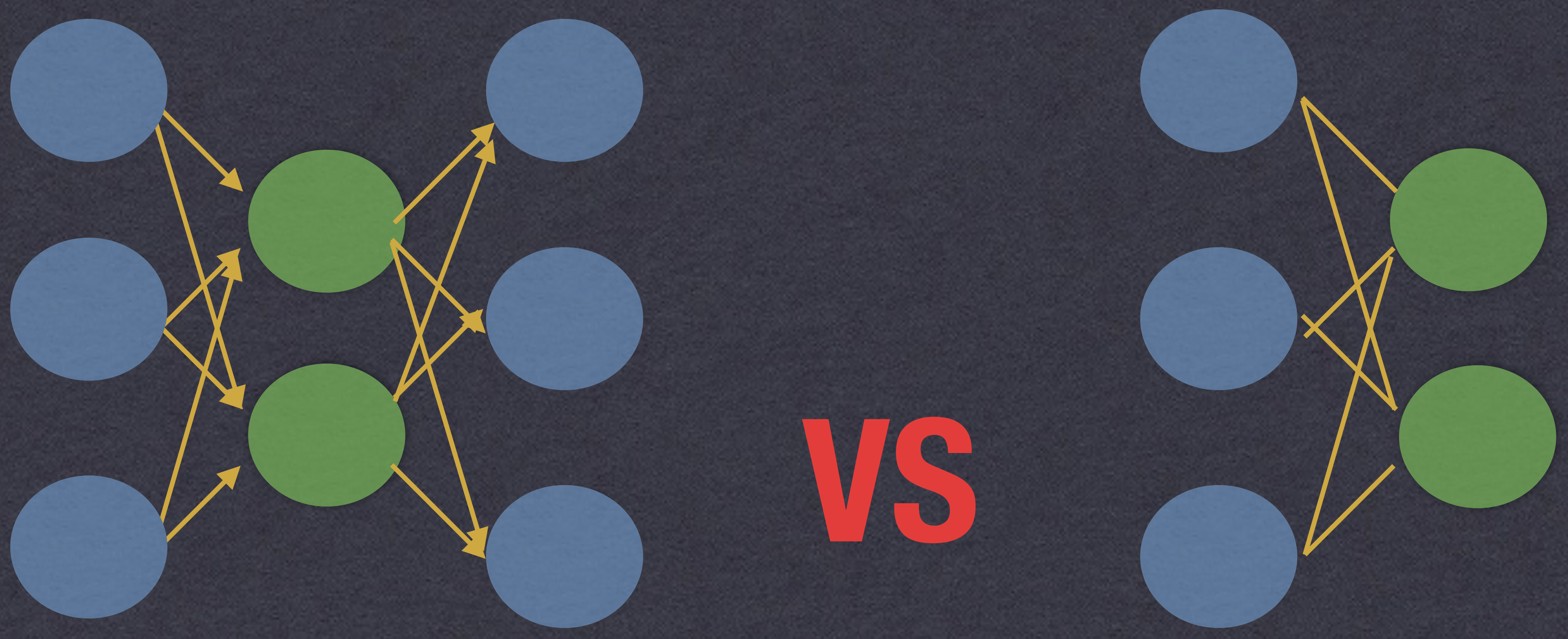
DANIEL KOHLSDORF

LAST LECTURE: DEEP AUTO ENCODERS



- Directed Model
- Reconstructs the input
- Back propagation
- Today:
  - Probabilistic Interpretation
  - Undirected Model

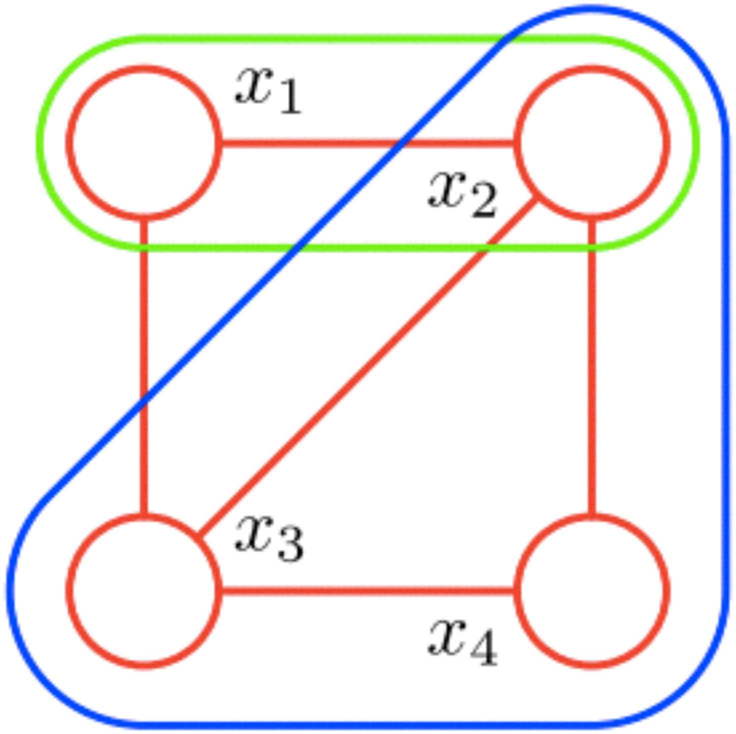
DIRECTED VS UNDIRECTED MODELS

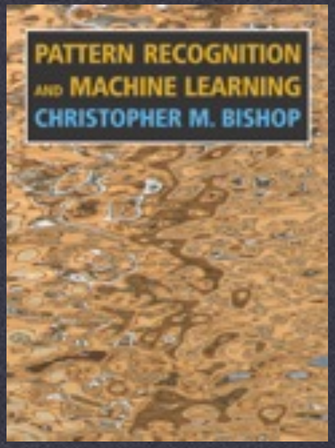


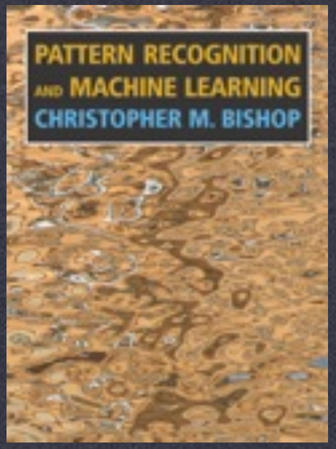
PROBABILISTIC UNDIRECTED MODELS

$$p(\mathbf{x}) = \frac{1}{Z} \prod_C \psi_C(\mathbf{x}_C).$$

$$Z = \sum_{\mathbf{x}} \prod_C \psi_C(\mathbf{x}_C)$$

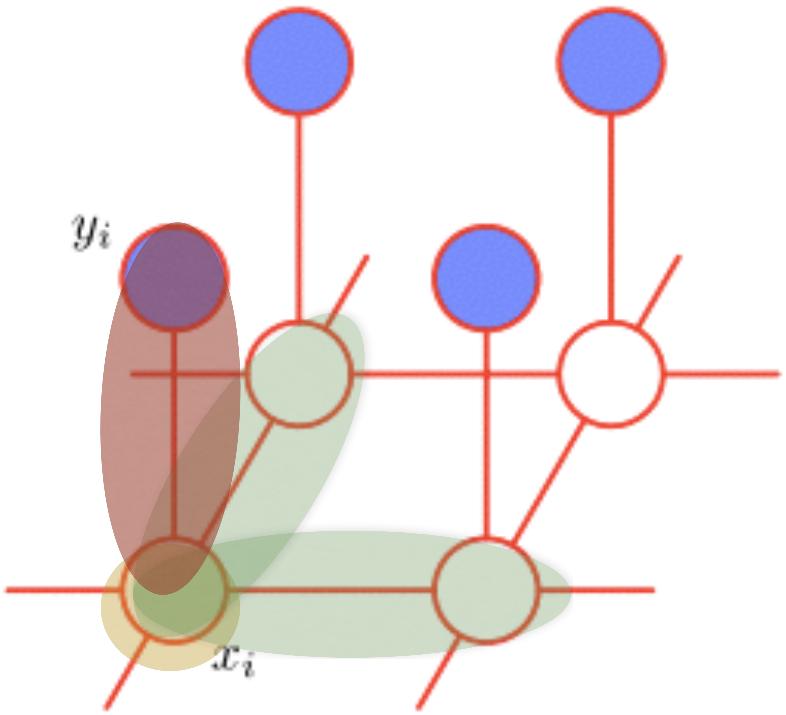






PRELIMINARIES: MARKOV RANDOM FIELD

Figure 8.31 An undirected graphical model representing a Markov random field for image de-noising, in which x_i is a binary variable denoting the state of pixel i in the unknown noise-free image, and y_i denotes the corresponding value of pixel i in the observed noisy image.

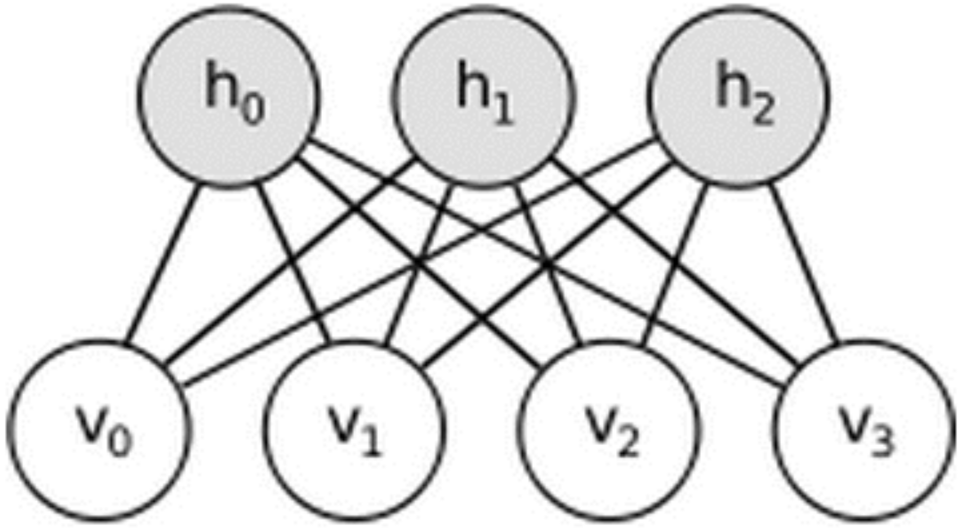


$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{Z} \exp\{-E(\mathbf{x}, \mathbf{y})\}$$

$$E(\mathbf{x}, \mathbf{y}) = h \sum_i x_i - \beta \sum_{\{i,j\}} x_i x_j - \eta \sum_i x_i y_i$$

Probability Distribution **Cliques**

RESTRICTED BOLTZMANN MACHINE

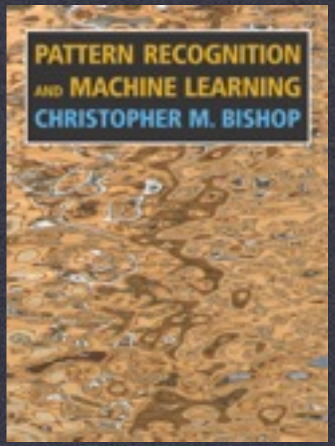


$$p(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} \, e^{-E(\mathbf{v},\mathbf{h})}$$

$$E(\mathbf{v}, \mathbf{h}) = - \sum_{i \in \text{visible}} a_i v_i - \sum_{j \in \text{hidden}} b_j h_j - \sum_{i,j} v_i h_j w_{ij}$$

Hinton: A Practical Guide to Training Restricted Boltzmann Machines

GIBBS SAMPLING



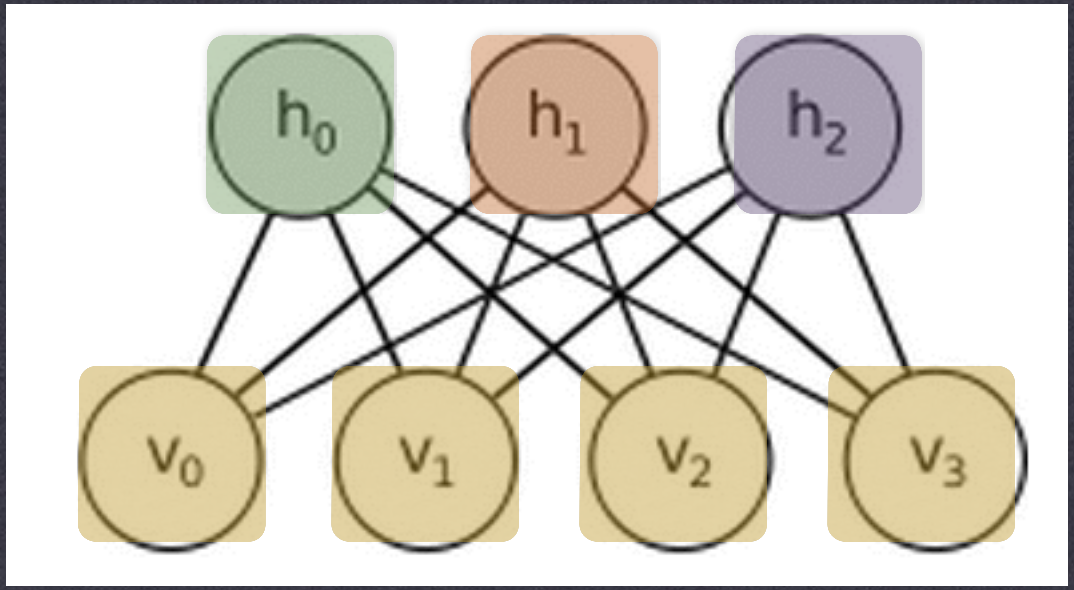
Gibbs Sampling

1. Initialize $\{z_i : i = 1, \ldots, M\}$
2. For $\tau = 1, \ldots, T$:
   - Sample $z_1^{(\tau+1)} \sim p(z_1 | z_2^{(\tau)}, z_3^{(\tau)}, \ldots, z_M^{(\tau)})$.
   - Sample $z_2^{(\tau+1)} \sim p(z_2 | z_1^{(\tau+1)}, z_3^{(\tau)}, \ldots, z_M^{(\tau)})$.

     $\vdots$
   - Sample $z_j^{(\tau+1)} \sim p(z_j | z_1^{(\tau+1)}, \ldots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \ldots, z_M^{(\tau)})$.

     $\vdots$
   - Sample $z_M^{(\tau+1)} \sim p(z_M | z_1^{(\tau+1)}, z_2^{(\tau+1)}, \ldots, z_{M-1}^{(\tau+1)})$.

GIBBS SAMPLING FOR RBM



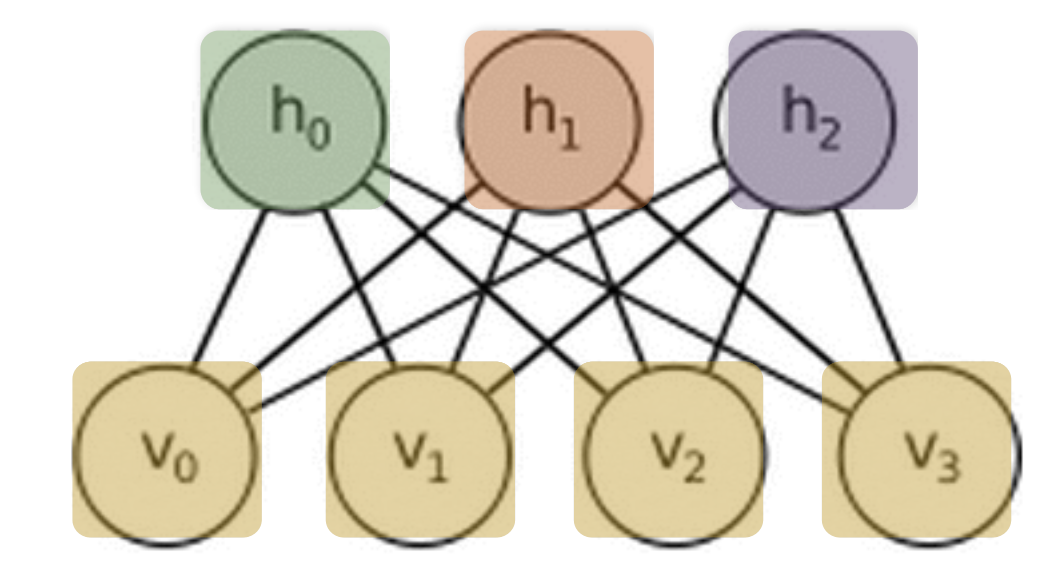
$h_0 \sim p(h_0 \mid v_0, v_1, v_2, v_3, h_1, h_2)$

$h_1 \sim p(h_1 \mid v_0, v_1, v_2, v_3, h_0, h_2)$

$h_2 \sim p(h_2 \mid v_0, v_1, v_2, v_3, h_1, h_0)$

h_0, h_1, h_2 are independent

GIBBS SAMPLING FOR RBM



$h_0 \sim p(h_0 \mid v_0, v_1, v_2, v_3)$

$h_1 \sim p(h_1 \mid v_0, v_1, v_2, v_3)$

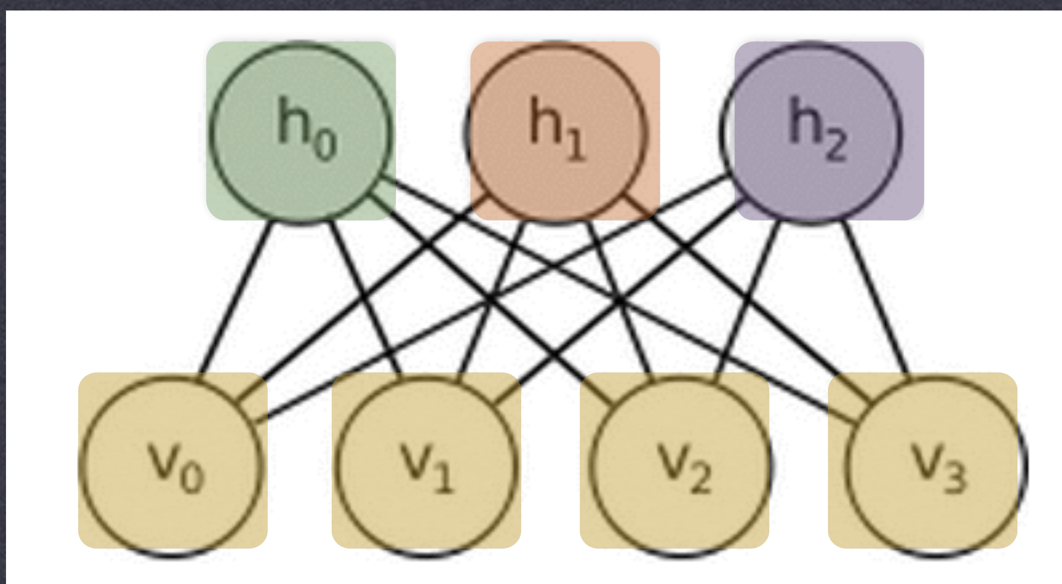
$h_2 \sim p(h_2 \mid v_0, v_1, v_2, v_3)$

RBM HIDDEN CONDITIONAL

$p(h_0 \mid v_0, v_1, v_2, v_3)$

$$p(h_j = 1 \mid \mathbf{v}) = \sigma(b_j + \sum_i v_i w_{ij})$$

RBM HIDDEN CONDITIONAL



Gibbs Sampling

1. Initialize $\{z_i : i = 1, \ldots, M\}$
2. For $\tau = 1, \ldots, T$:
   - Sample $z_1^{(\tau+1)} \sim p(z_1 | z_2^{(\tau)}, z_3^{(\tau)}, \ldots, z_M^{(\tau)})$.
   - Sample $z_2^{(\tau+1)} \sim p(z_2 | z_1^{(\tau+1)}, z_3^{(\tau)}, \ldots, z_M^{(\tau)})$.

     $\vdots$
   - Sample $z_j^{(\tau+1)} \sim p(z_j | z_1^{(\tau+1)}, \ldots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \ldots, z_M^{(\tau)})$.

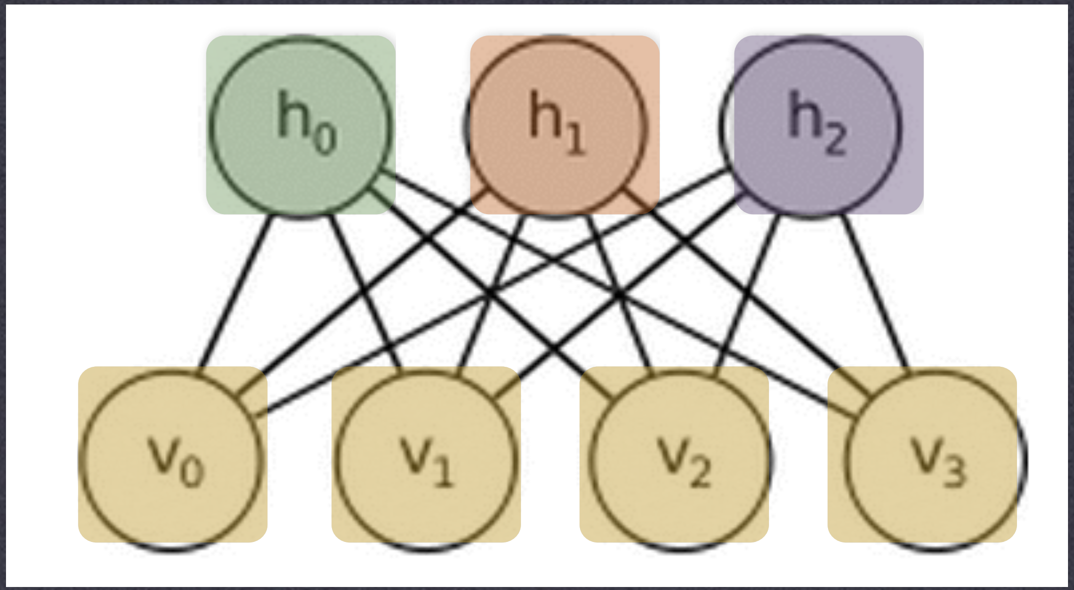
     $\vdots$
   - Sample $z_M^{(\tau+1)} \sim p(z_M | z_1^{(\tau+1)}, z_2^{(\tau+1)}, \ldots, z_{M-1}^{(\tau+1)})$.

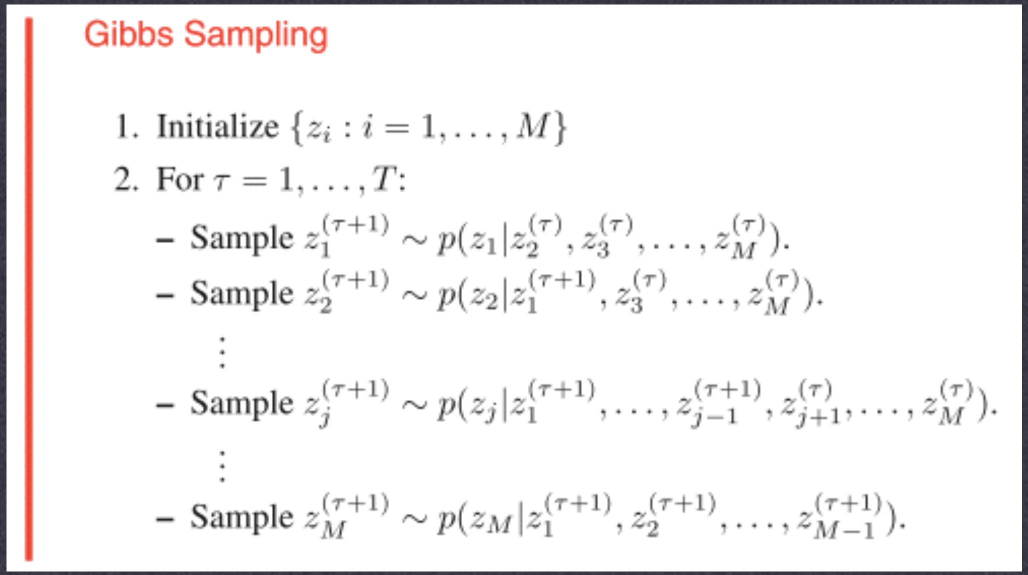
$$h_0 = p(h_0 = 1 \mid v_0, v_1, v_2, v_3) > \mathrm{Uniform}(0, 1)$$

$$h_1 = p(h_1 = 1 \mid v_0, v_1, v_2, v_3) > \mathrm{Uniform}(0, 1)$$

$$h_2 = p(h_2 = 1 \mid v_0, v_1, v_2, v_3) > \mathrm{Uniform}(0, 1)$$

RBM VISIBLE CONDITIONAL





Gibbs Sampling

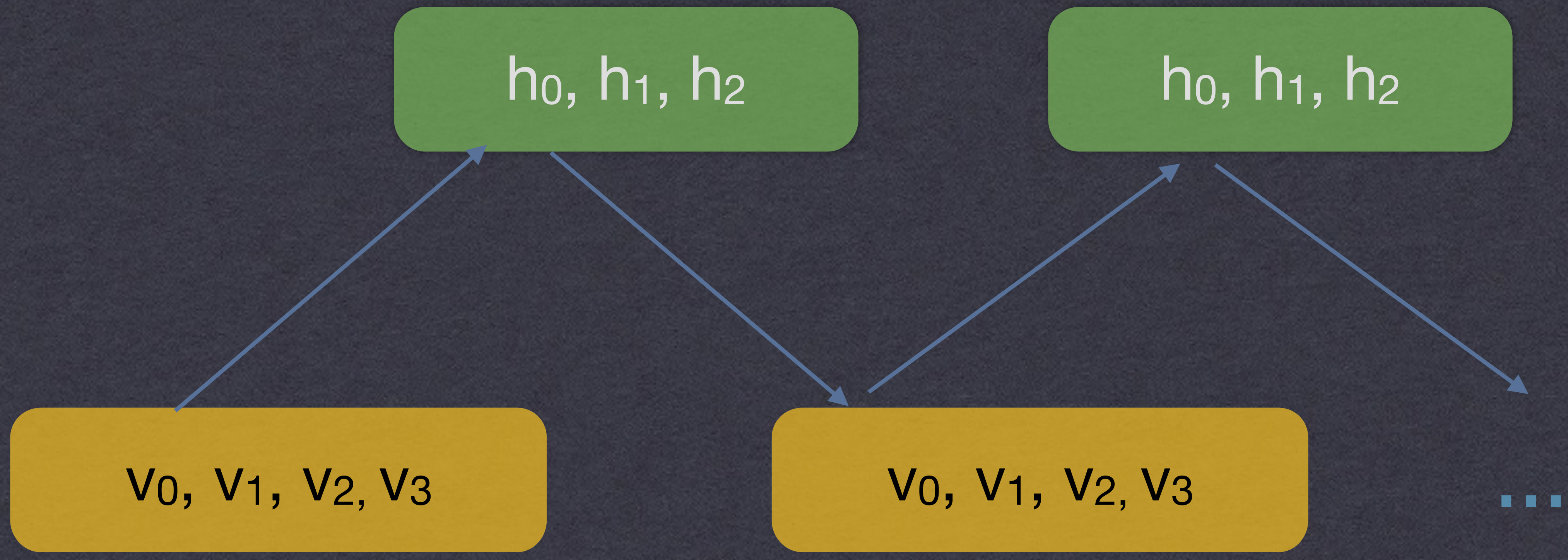
1. Initialize $\{z_i : i = 1, \ldots, M\}$
2. For $\tau = 1, \ldots, T$:
   - Sample $z_1^{(\tau+1)} \sim p(z_1 | z_2^{(\tau)}, z_3^{(\tau)}, \ldots, z_M^{(\tau)})$.
   - Sample $z_2^{(\tau+1)} \sim p(z_2 | z_1^{(\tau+1)}, z_3^{(\tau)}, \ldots, z_M^{(\tau)})$.

     $\vdots$
   - Sample $z_j^{(\tau+1)} \sim p(z_j | z_1^{(\tau+1)}, \ldots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \ldots, z_M^{(\tau)})$.

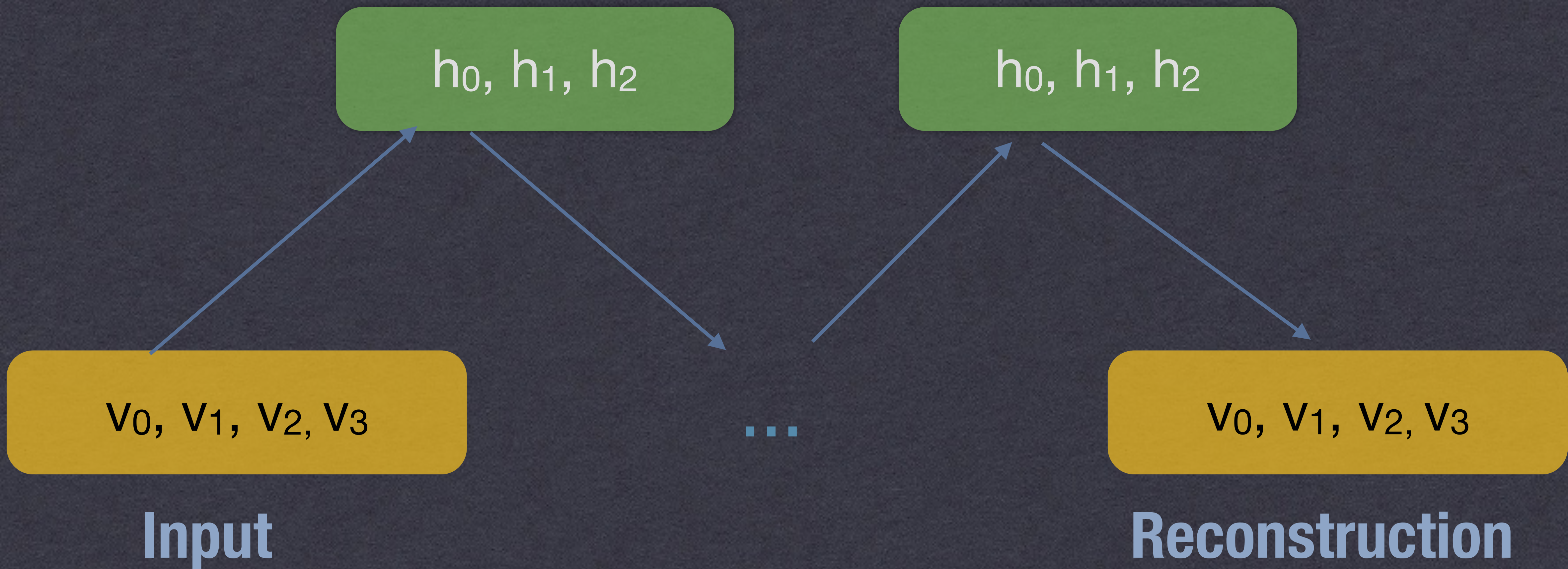
     $\vdots$
   - Sample $z_M^{(\tau+1)} \sim p(z_M | z_1^{(\tau+1)}, z_2^{(\tau+1)}, \ldots, z_{M-1}^{(\tau+1)})$.

$$v_0 = p(v_0 = 1 \mid h_0, h_1, h_2) > \text{Uniform}(0, 1)$$
$$v_1 = p(v_1 = 1 \mid h_0, h_1, h_2) > \text{Uniform}(0, 1)$$
$$v_2 = p(v_2 = 1 \mid h_0, h_1, h_2) > \text{Uniform}(0, 1)$$
$$v_3 = p(v_3 = 1 \mid h_0, h_1, h_2) > \text{Uniform}(0, 1)$$

ALTERNATE GIBBS SAMPLING



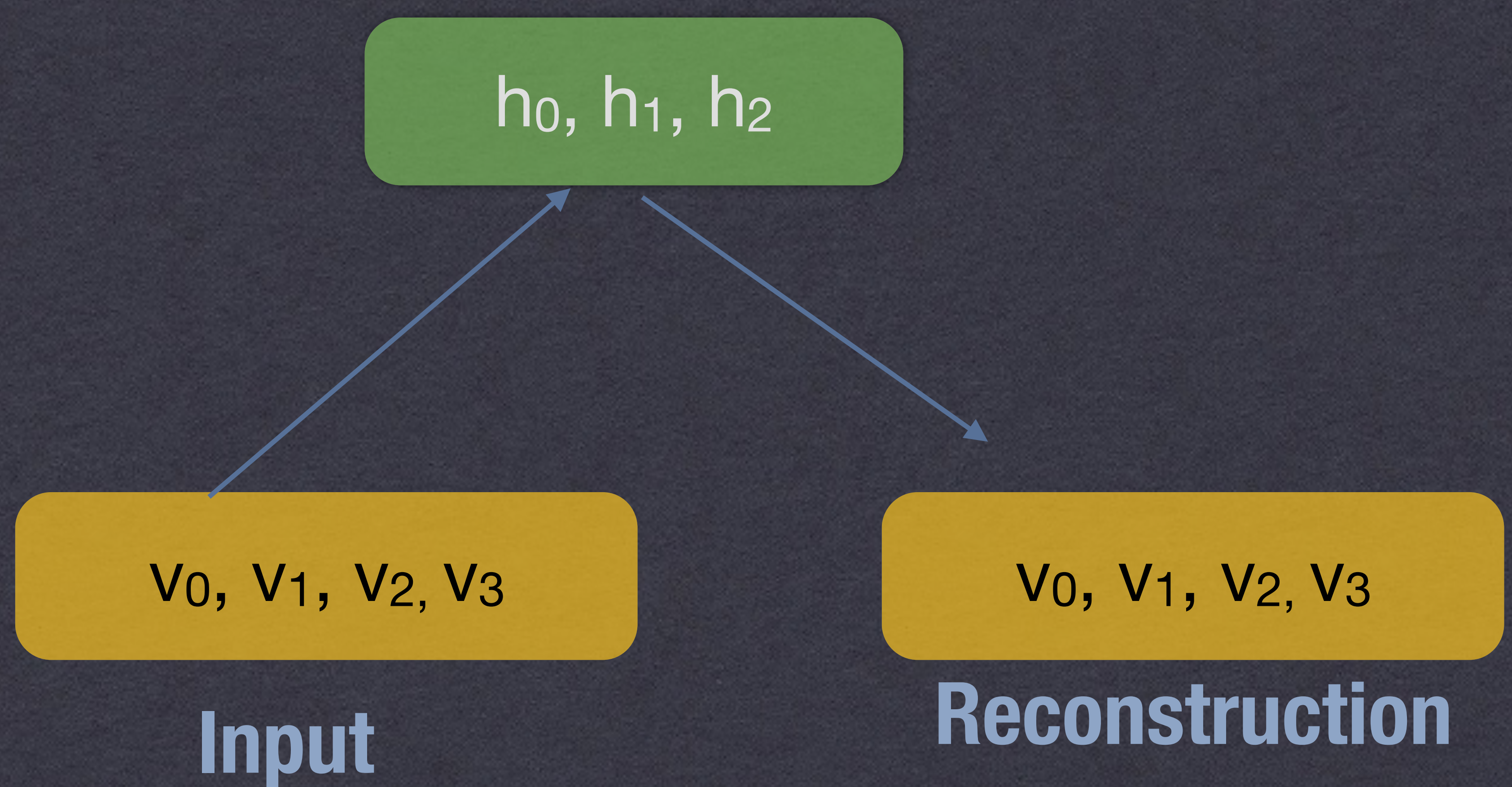
LEARNING



Weight Update

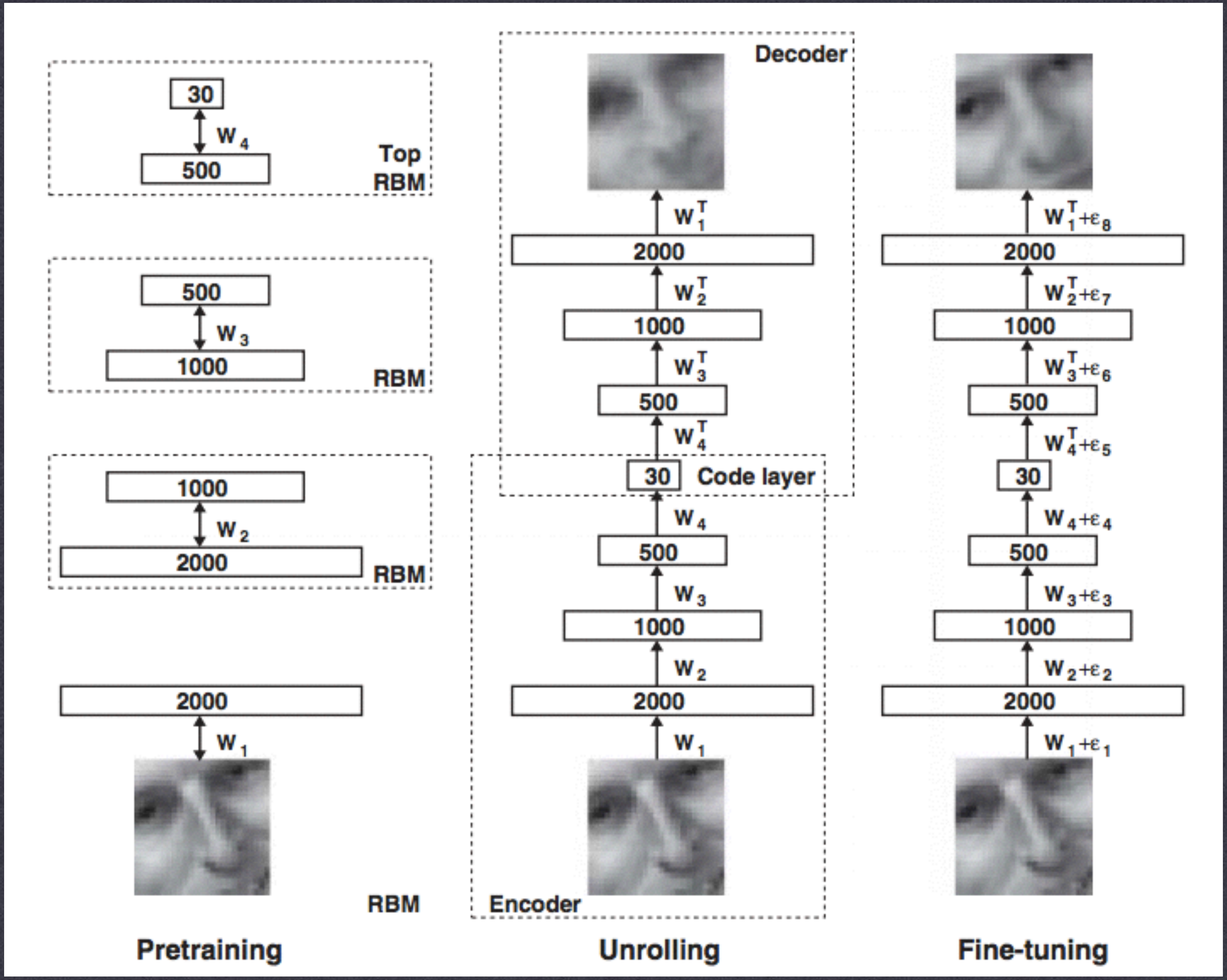
LEARNING: CONTRASTIVE DIVERGENCE

Just do it once!

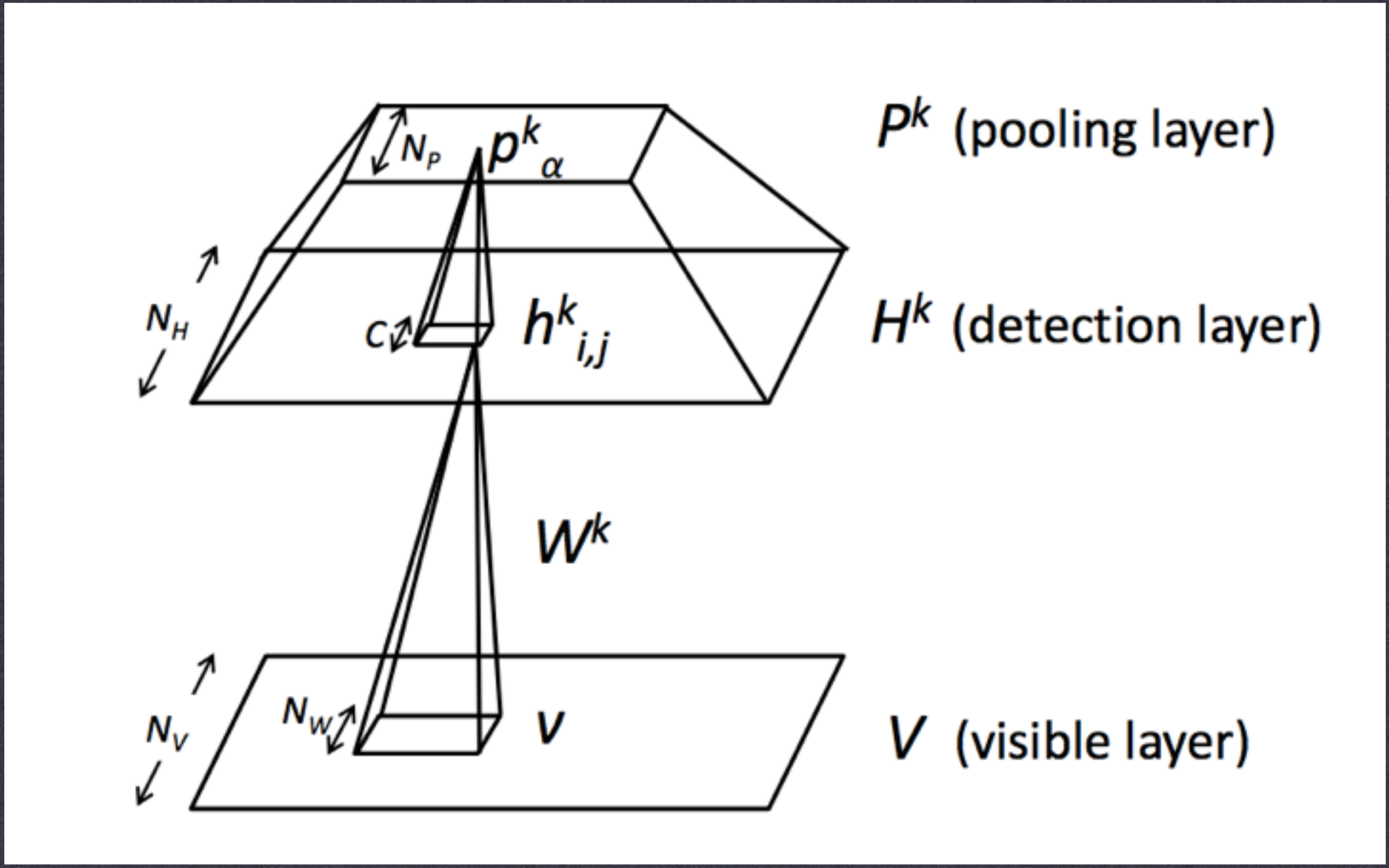


DEEP!

Hinton, G. E. and Salakhutdinov, R. R. (2006)
Reducing the dimensionality of data with neural networks.
Science



CONVOLUTIONAL RESTRICTED BOLTZMANN MACHINES



Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng.

- Convolution:
  - Flip the filter in both dimensions (right to left, bottom to top)
  - Then apply cross-correlation

$$G[i,j] = \sum_{u=-k}^{k} \sum_{v=-k}^{k} H[u,v]F[i-u, j-v]$$

$$G = H * F$$

Notation for convolution operator

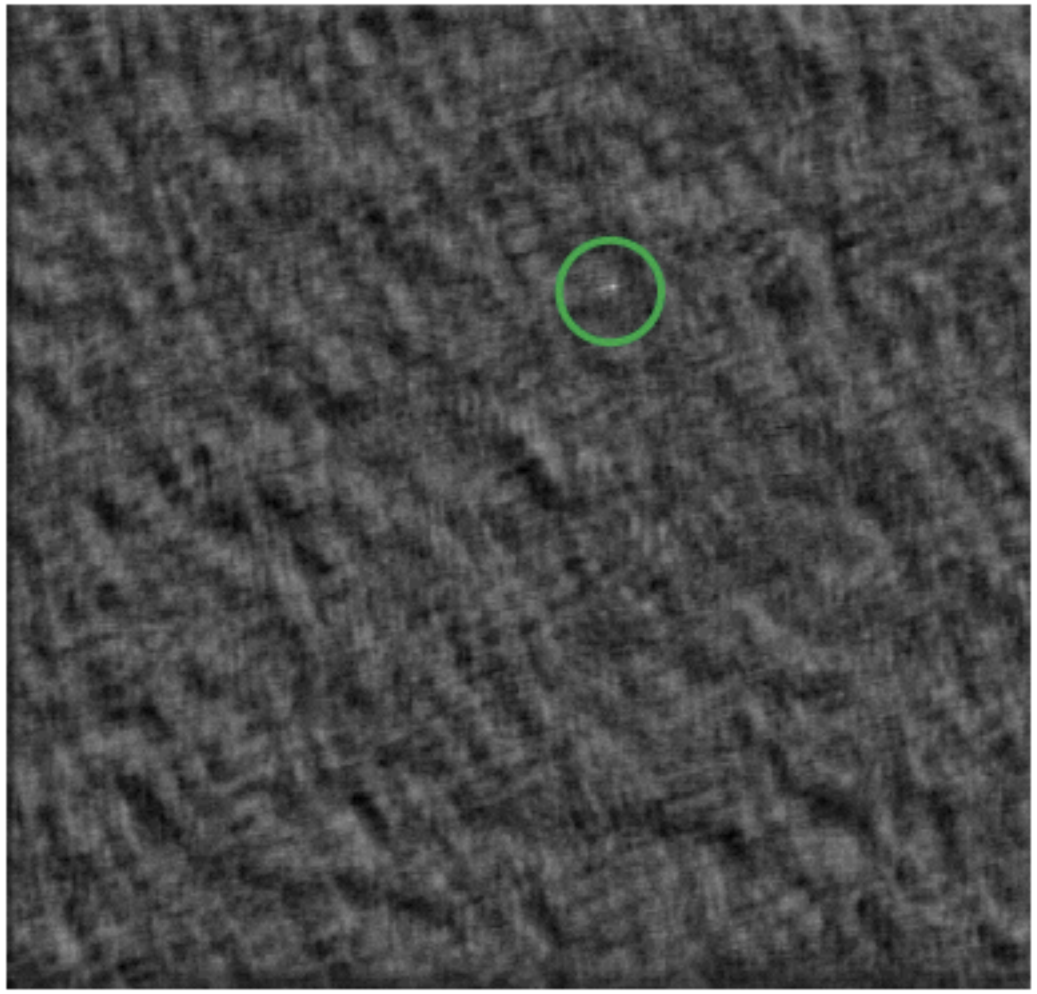
K. Grauman

Edge Detector

Gaussian

From Aaron

Detected template

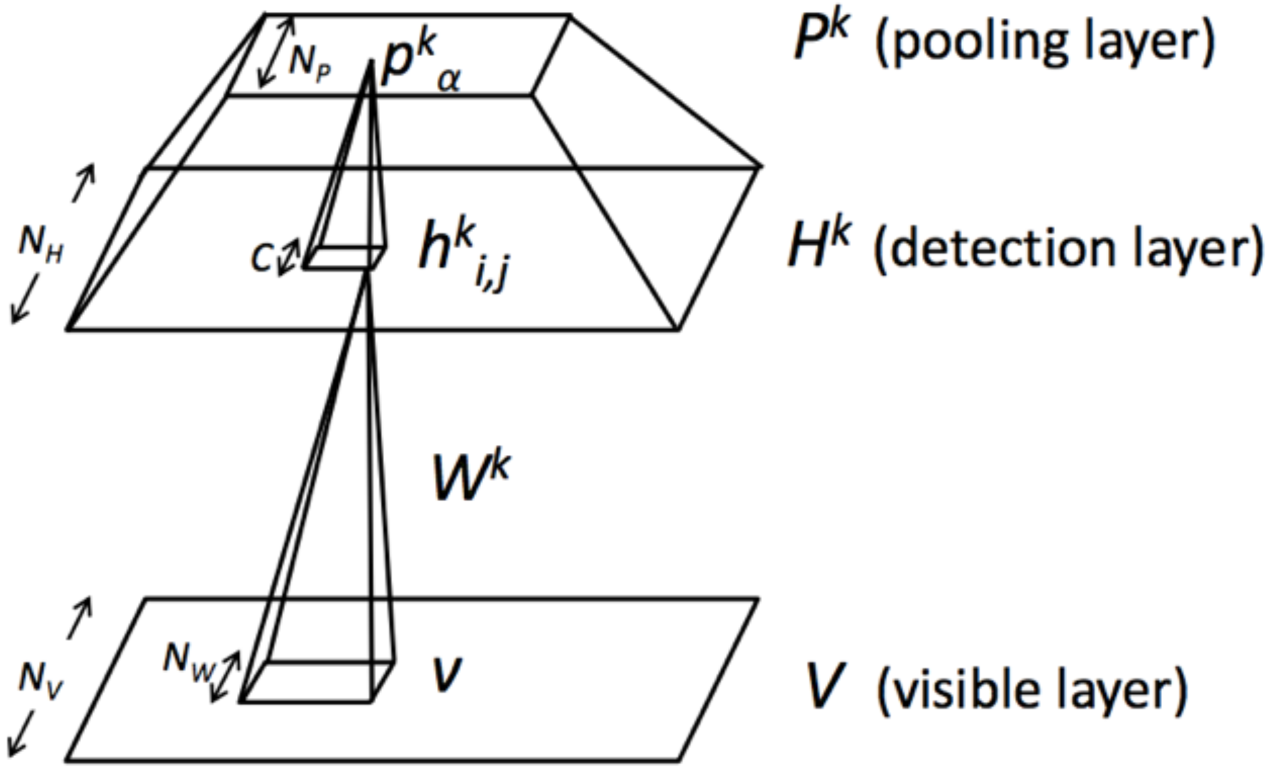


Correlation map

Template

From Aaron

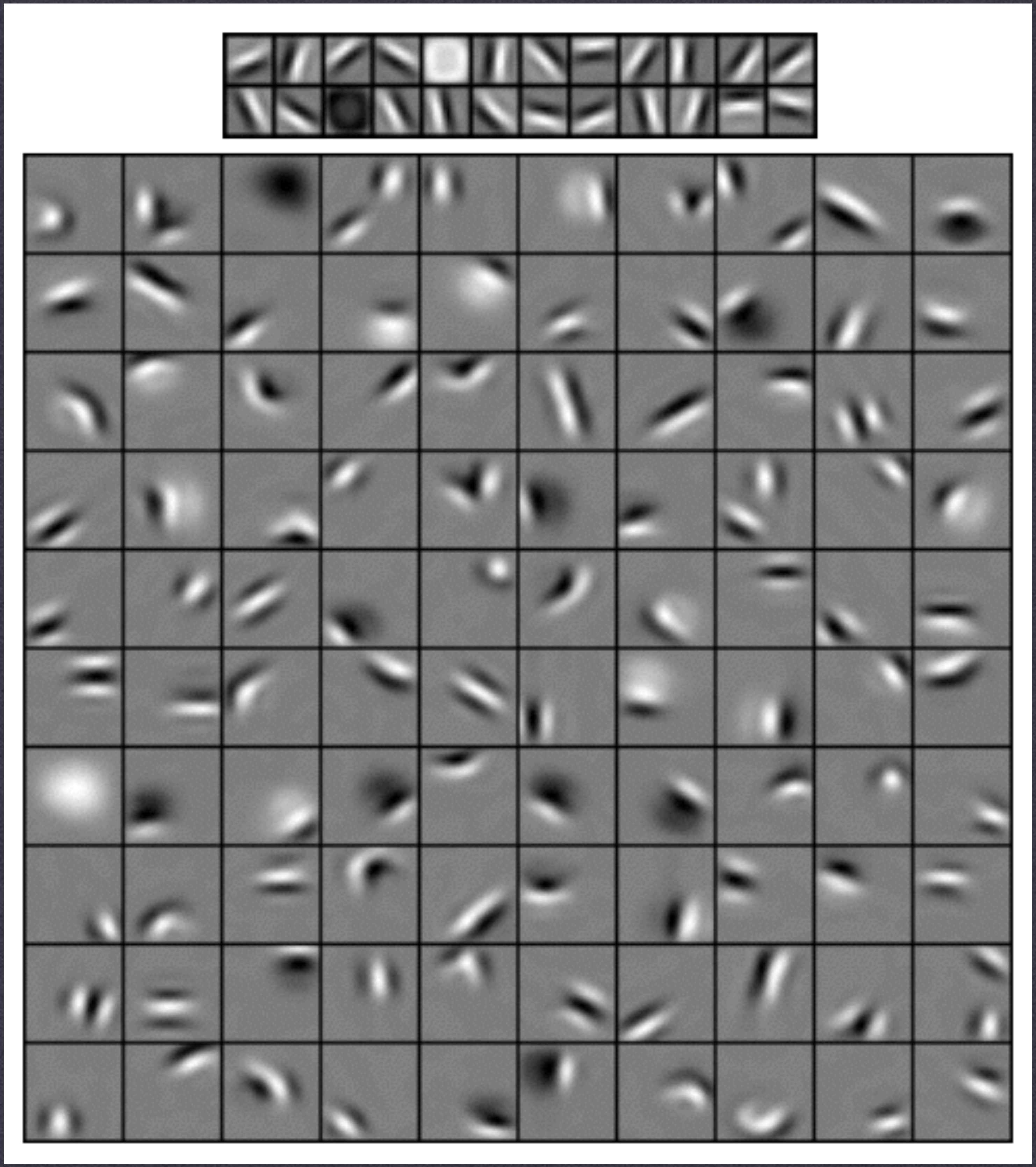
Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng.



$$P(h^k_{ij} = 1|\mathbf{v}) = \sigma((\tilde{W}^k * v)_{ij} + b_k)$$

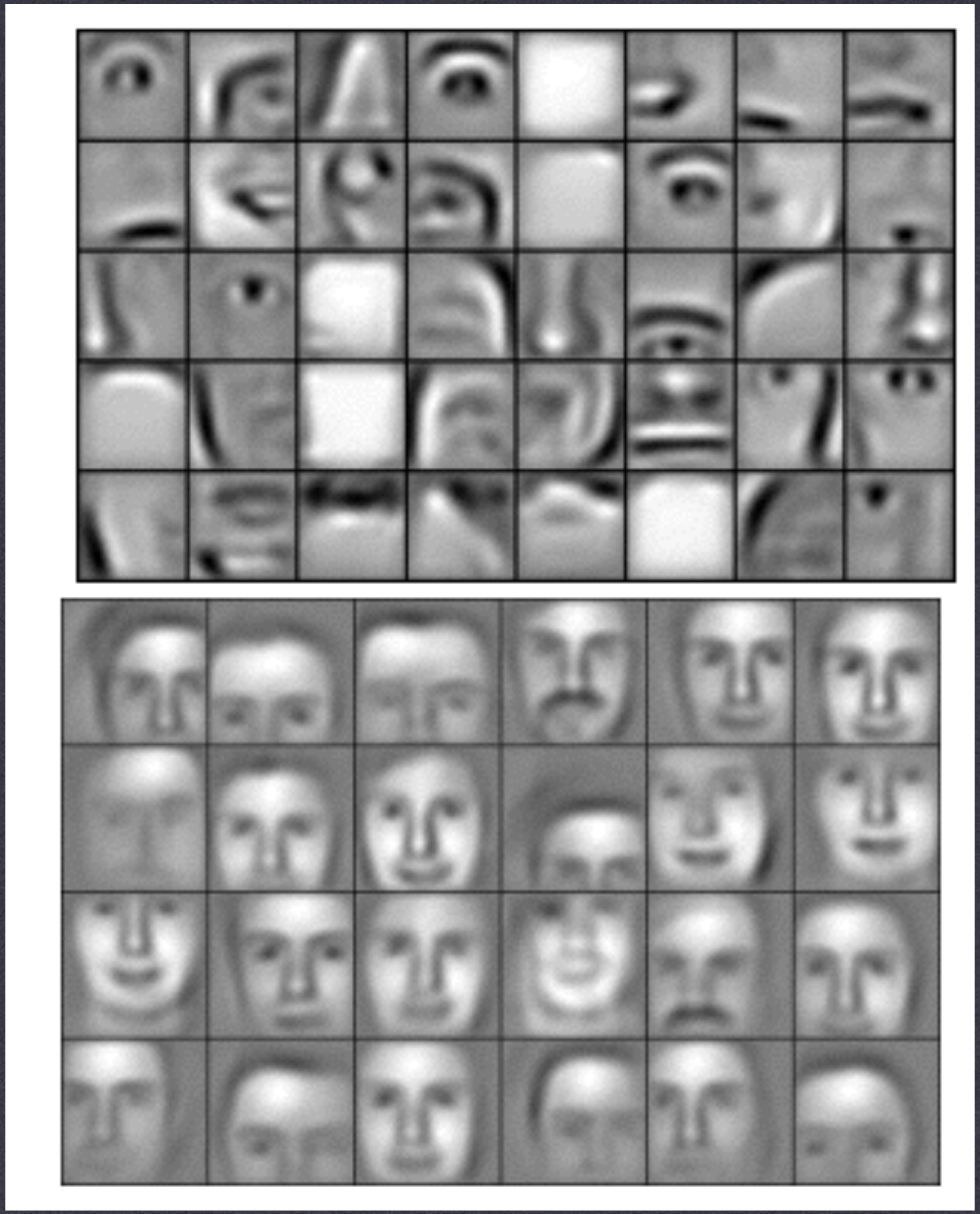
$$P(v_{ij} = 1|\mathbf{h}) = \sigma((\sum_k W^k * h^k)_{ij} + c)$$

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{k=1}^{K} h^k \bullet (\tilde{W}^k * v) - \sum_{k=1}^{K} b_k \sum_{i,j} h^k_{i,j} - c \sum_{i,j} v_{ij}$$

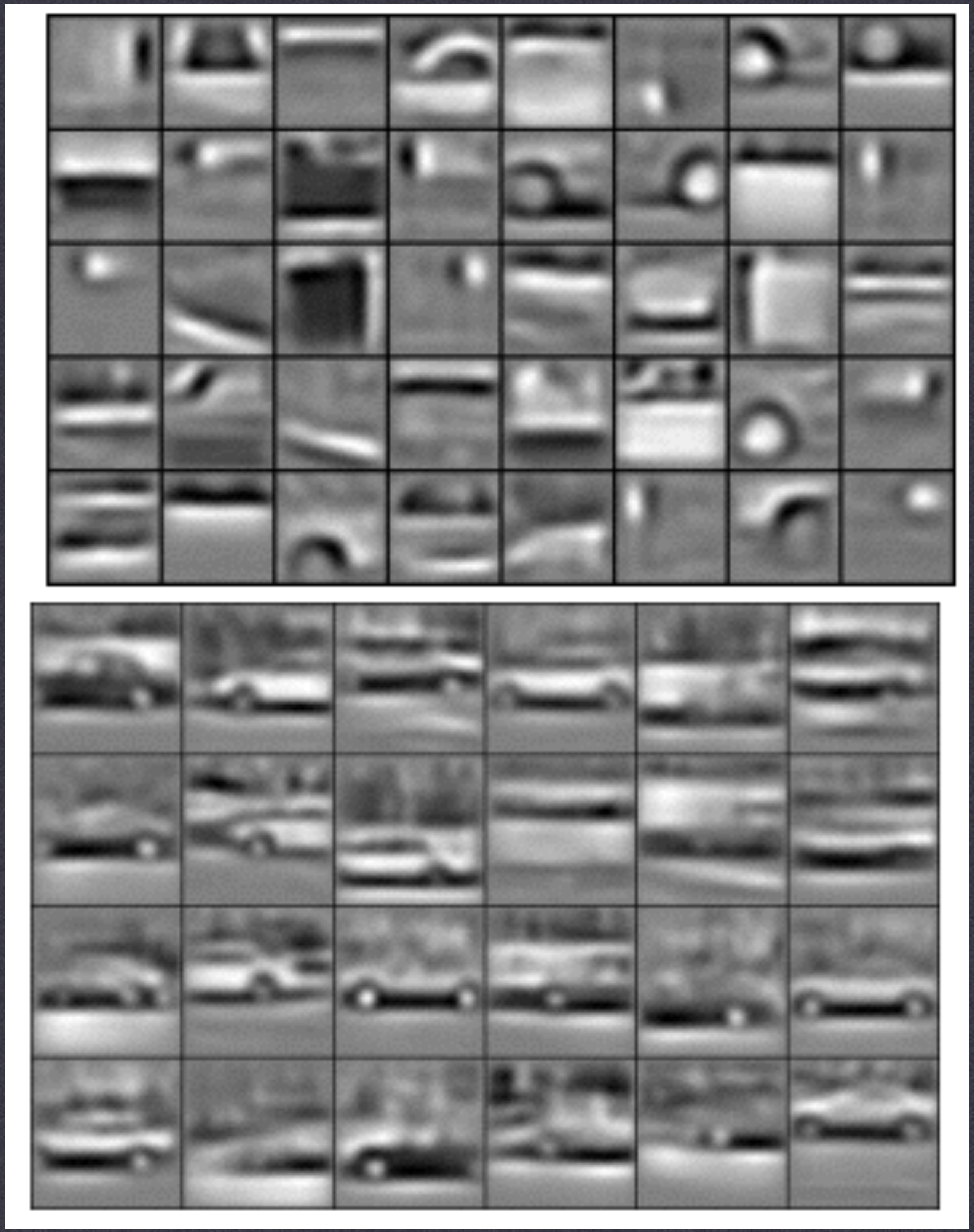


Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng.

Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng.



Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng.



Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng.