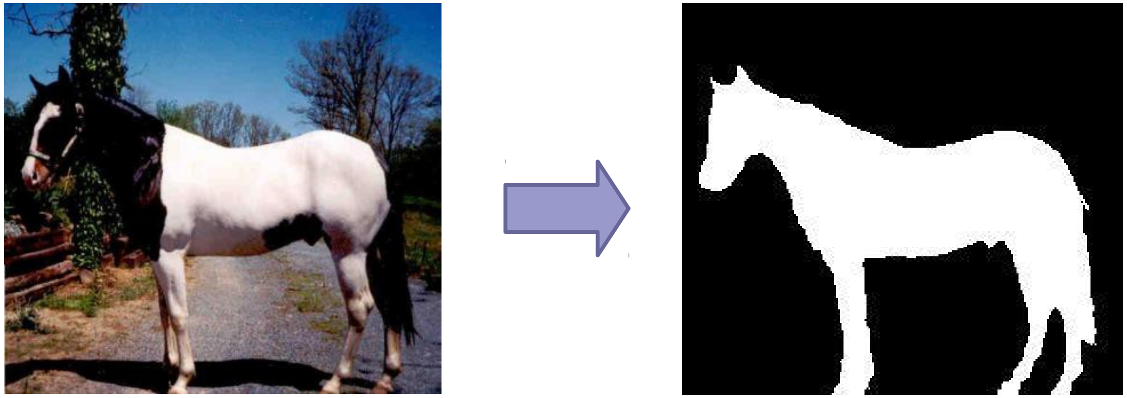
Exploring Compositional High Order Pattern Potentials for Structured Output Learning

Yujia Li, Daniel Tarlow*, Richard Zemel

University of Toronto
*Now at Microsoft Research Cambridge
June 25, 2013

Structured Output Learning

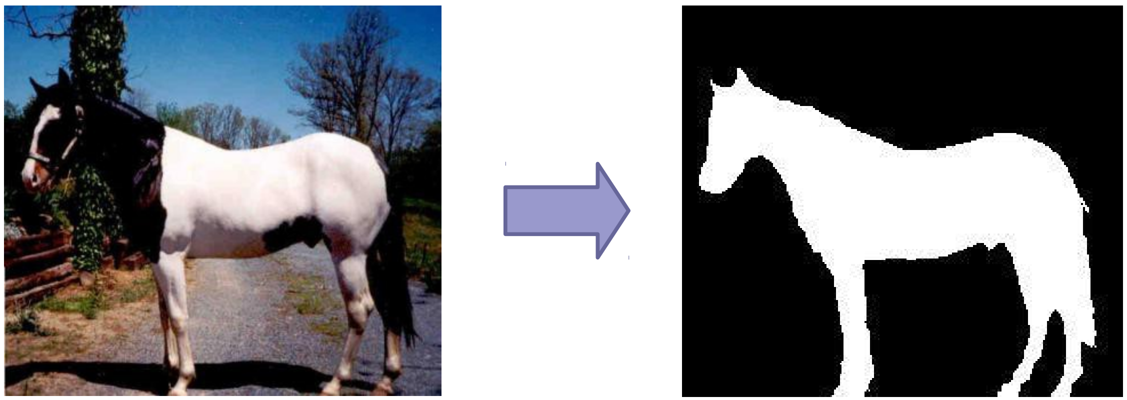
- Lots of real world applications require structured outputs
  - Image segmentation, pose estimation, sequence labeling, etc.



Figures from Weizmann horse dataset

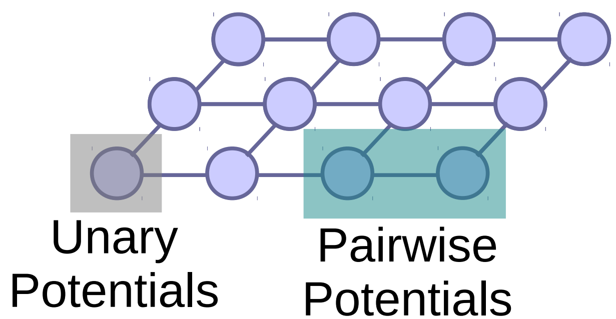
Structured Output Learning

- Lots of real world applications require structured outputs
  - Image segmentation, pose estimation, sequence labeling, etc.



- Standard model – pairwise MRF/CRF

$$E(\mathbf{y}) = \sum_i f_i^u(y_i) + \sum_{i,j} f_{ij}^p(y_i, y_j)$$



  - Sparse connections – easier to learn and do inference
  - Overly simplistic – only modeling up to 2nd order correlation in outputs

Figures from Weizmann horse dataset

Moving to More Expressive Models

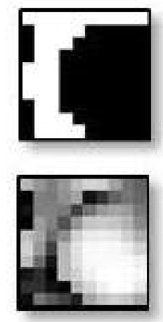
- Densely connected CRFs [P. Krahenbuhl et al. NIPS'12]
  - Still 2nd order connections but densely connected
- Robust High Order Potentials [P. Kohli et al. CVPR'08]
  - Smoothness in a region
- Global Connectivity Potentials [S. Nowozin et al. CVPR'09]
  - Require the output to be connected
- Pattern Potentials [C. Rother et al. CVPR'09]
  - Consistency between the output and learned patterns

Pattern Potentials

- Penalize linearly if output deviates from a pattern

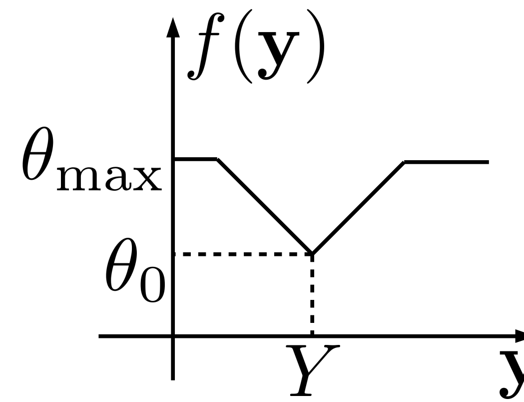
$$d(\mathbf{y}) = \sum_i abs(w_i)\mathbf{I}[y_i \neq Y_i]$$

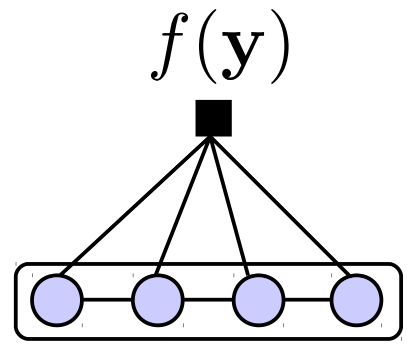
$$f(\mathbf{y}) = \min\{d(\mathbf{y}) + \theta_0, \theta_{\max}\}$$



Patterns

Weights





Pairwise CRF

- Multiple base pattern potentials can be combined to form more expressive composite pattern potentials

Pattern and weight figures: C. Rother et al. CVPR'09

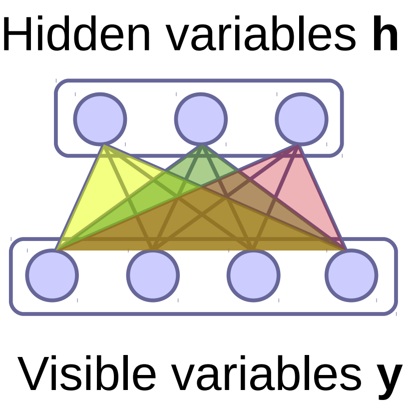
Restricted Boltzmann Machines (RBMs)

- RBM probabilistic model

$$E(\mathbf{y},\mathbf{h}) = -\sum_{ij} w_{ij} y_i h_j - \sum_i b_i y_i - \sum_j c_j h_j$$

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

  - Sum out **h**, RBM becomes a high order potential on **y**



- Some success modeling object shape
  - The Shape Boltzmann Machine [S. M. Ali Eslami et al., CVPR'12]
  - Masked RBMs [N. Heess et al. ICANN'11]

CHOPP

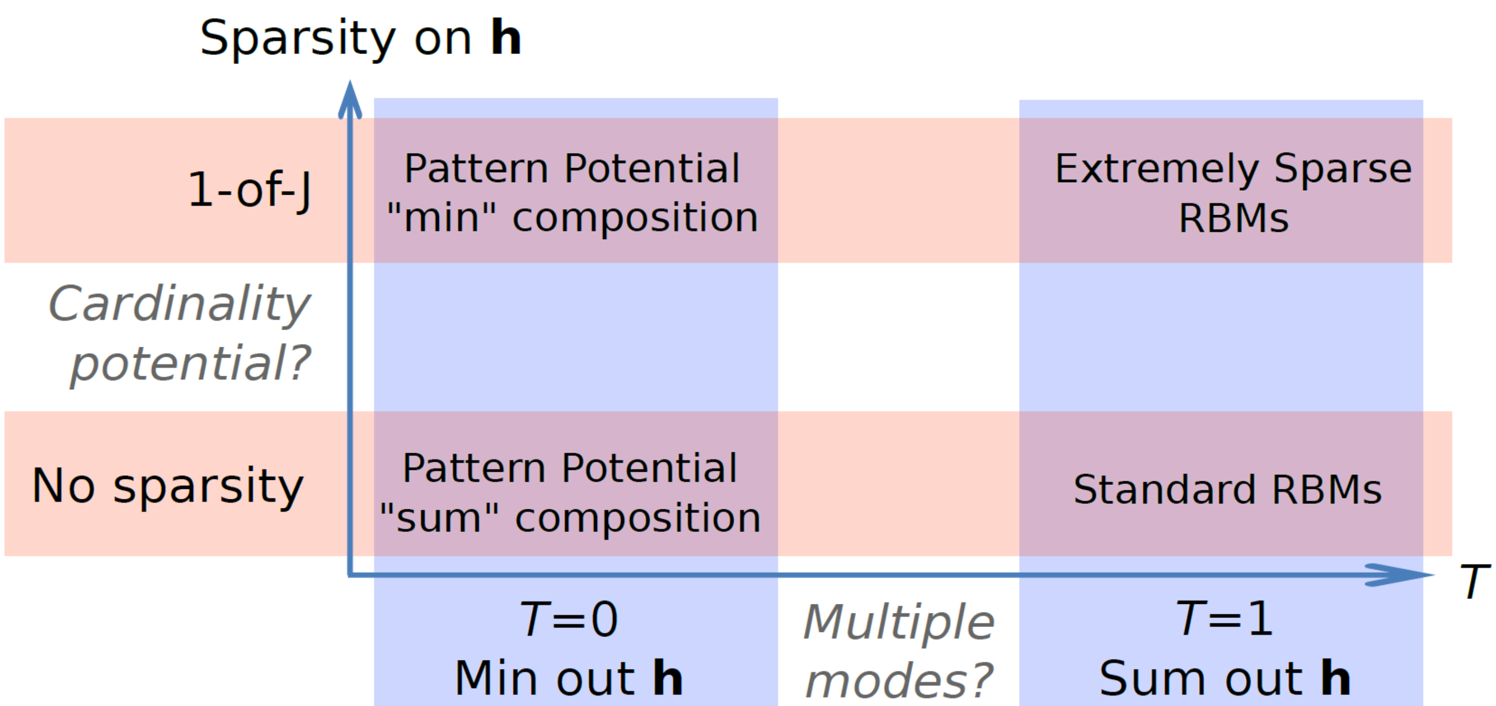
- Compositional High Order Pattern Potential (CHOPP)

$$f_T(\mathbf{y}) = -\boxed{T} \log \left(\sum_{\mathbf{h}} \exp \left(\frac{1}{T} \boxed{\sum_j} \left(c_j + \boxed{\sum_i w_{ij} y_i} \right) h_j \right) \right)$$

Interpolate between RBMs and PPs

Combine all patterns

Compatibility with a pattern



CHOPP-Augmented CRF

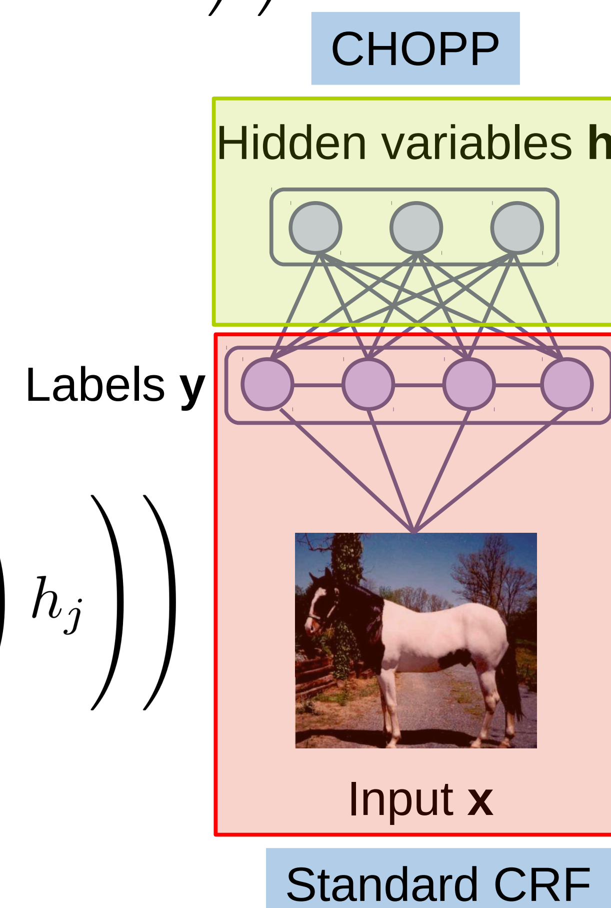
- Compositional High Order Pattern Potential (CHOPP)

$$f_T(\mathbf{y}) = -T \log \left(\sum_{\mathbf{h}} \exp \left(\frac{1}{T} \sum_j \left(c_j + \sum_i w_{ij} y_i \right) h_j \right) \right)$$

- CHOPP-augmented CRF Energy function

$$-E(\mathbf{y}|\mathbf{x}) = f^u(\mathbf{y}|\mathbf{x}) + f^p(\mathbf{y}|\mathbf{x})$$

$$+ T \log \left(\sum_{\mathbf{h}} \exp \left(\frac{1}{T} \sum_j \left(c_j + \sum_i w_{ij} y_j \right) h_j \right) \right)$$

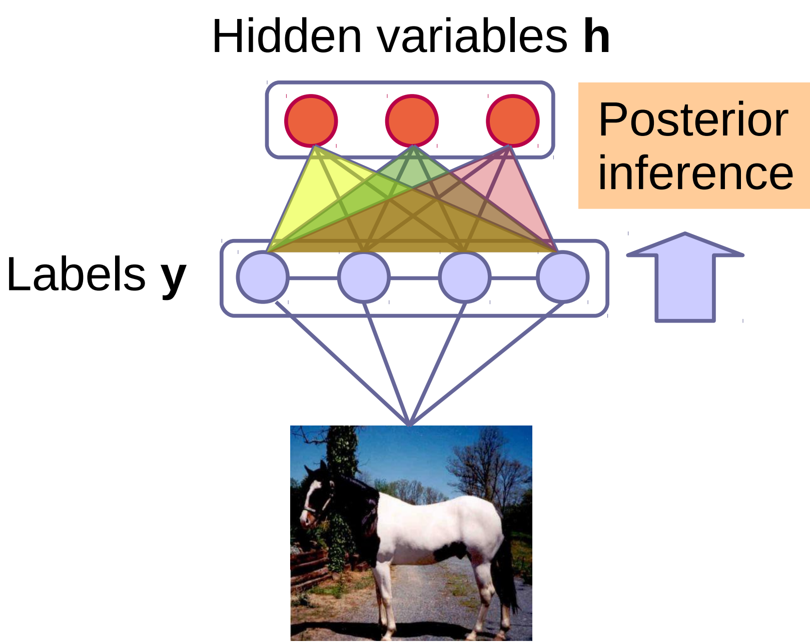


“EM” Inference Algorithm

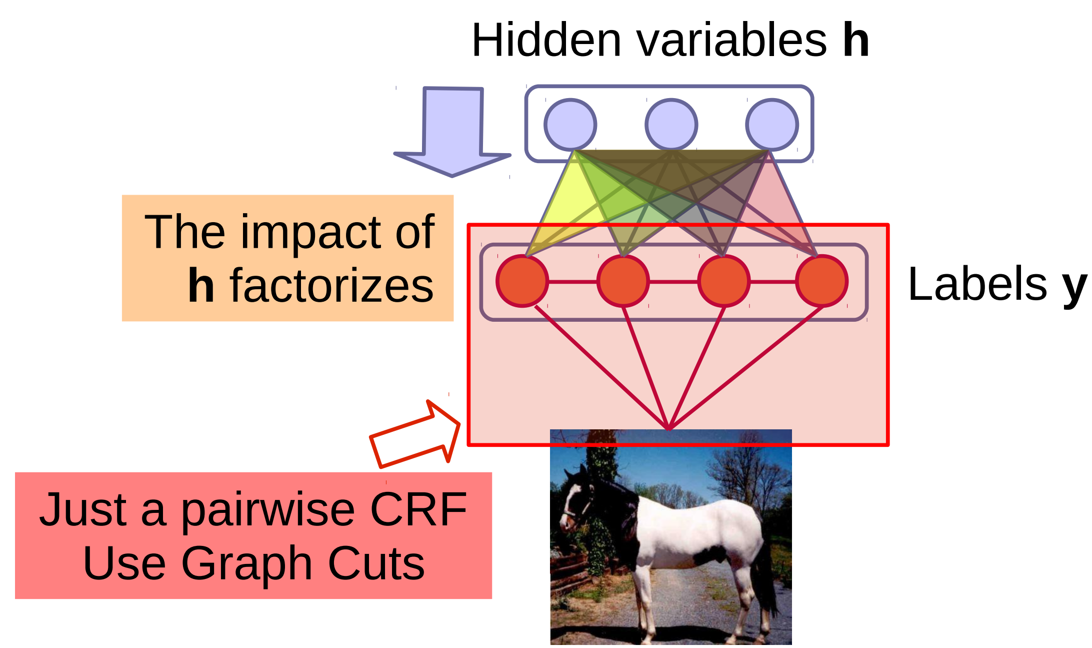
- Making predictions

$$\mathbf{y}^* = \underset{\mathbf{y}}{\operatorname{argmax}} \{-E(\mathbf{y}|\mathbf{x})\}$$

E-step: fix **y** compute **h**



M-step: fix **h** find optimal **y**



An Example for the “EM” Inference Algorithm

Original Image

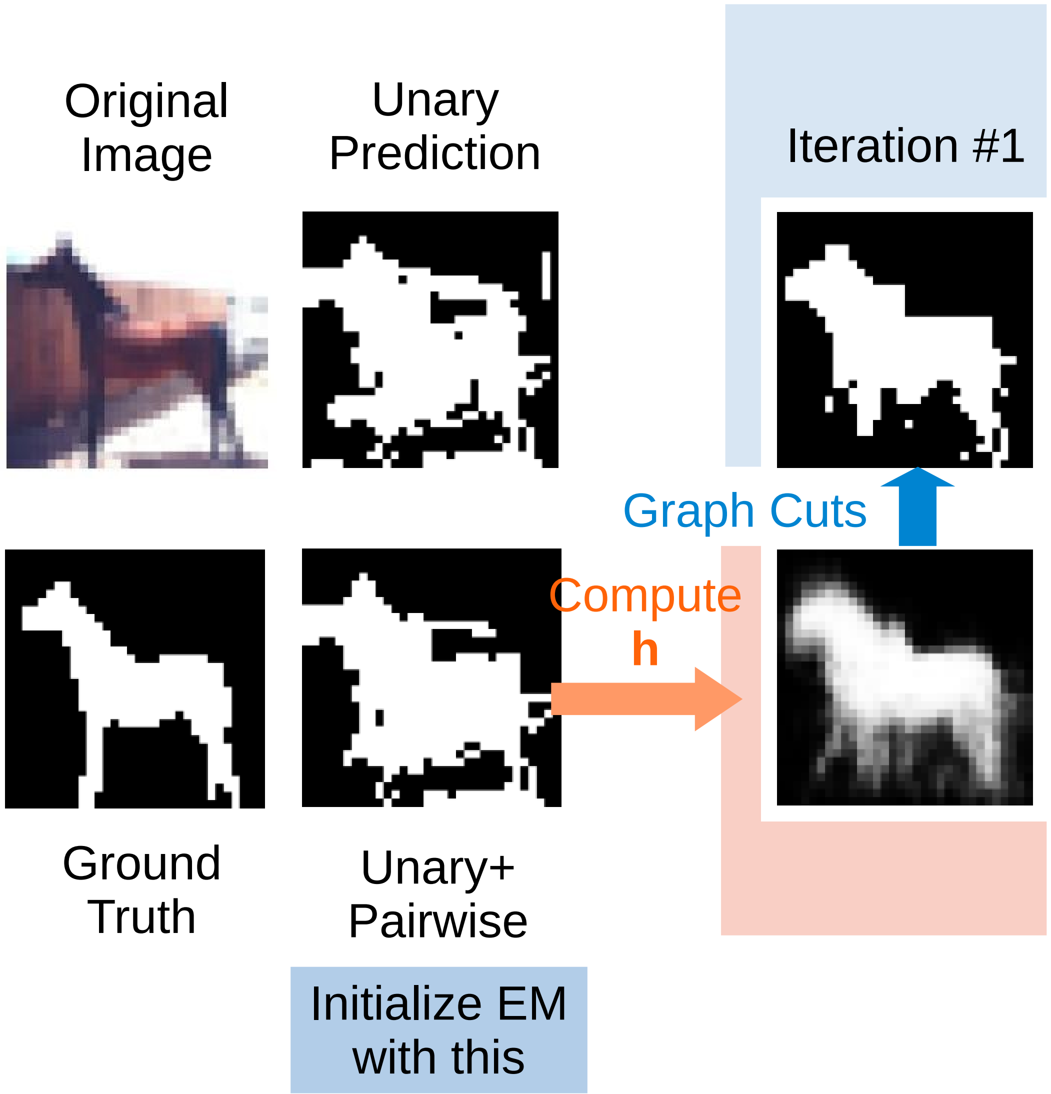
Unary Prediction

Ground Truth

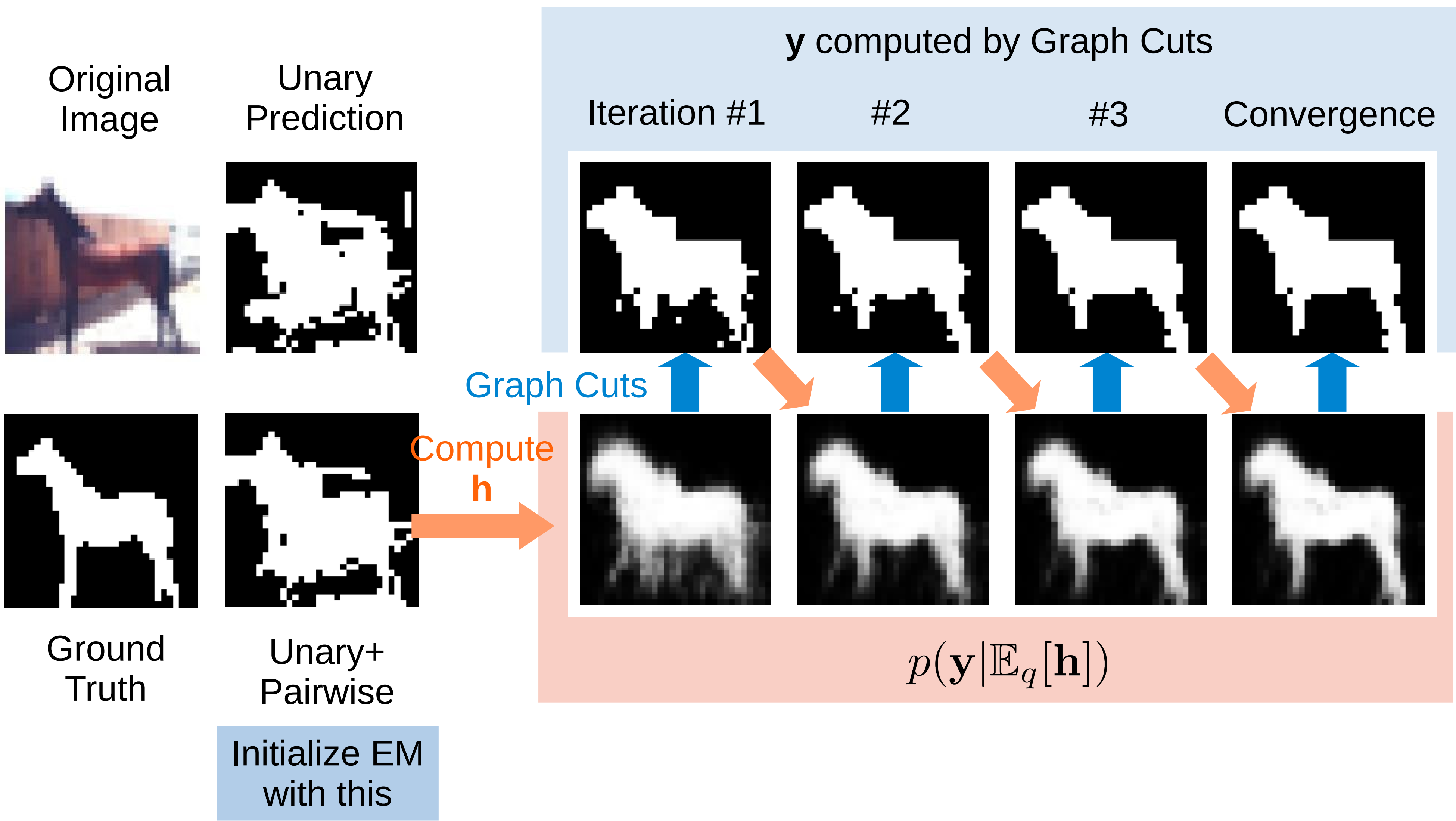
Unary+ Pairwise

Initialize EM with this

An Example for the “EM” Inference Algorithm



An Example for the “EM” Inference Algorithm

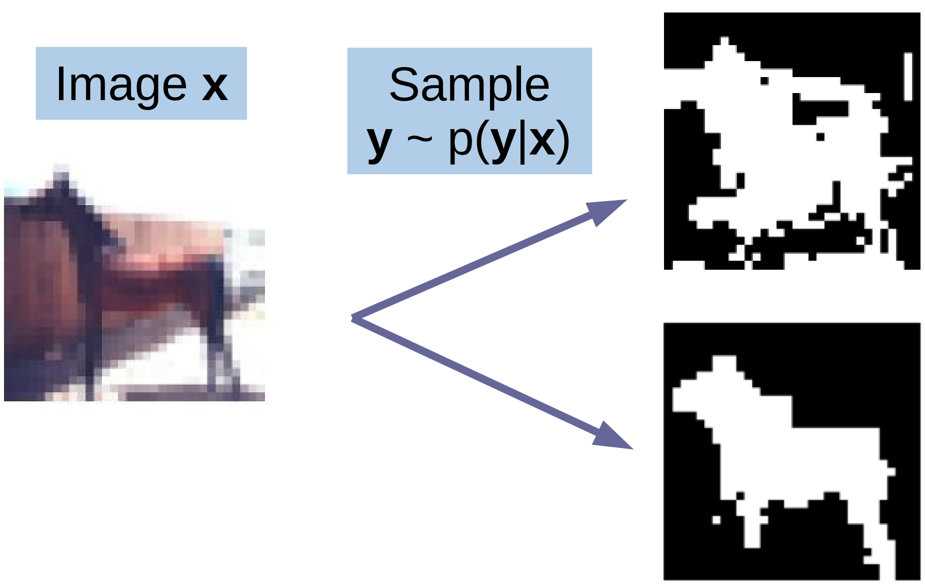


Learning by Minimizing Expected Loss

- Contrastive Divergence does not work well
- Expected loss objective

$$L = \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x};\theta)\ell(\mathbf{y},\mathbf{y}^*)$$

- Estimate gradient using a set of samples from $p(\mathbf{y}|\mathbf{x})$

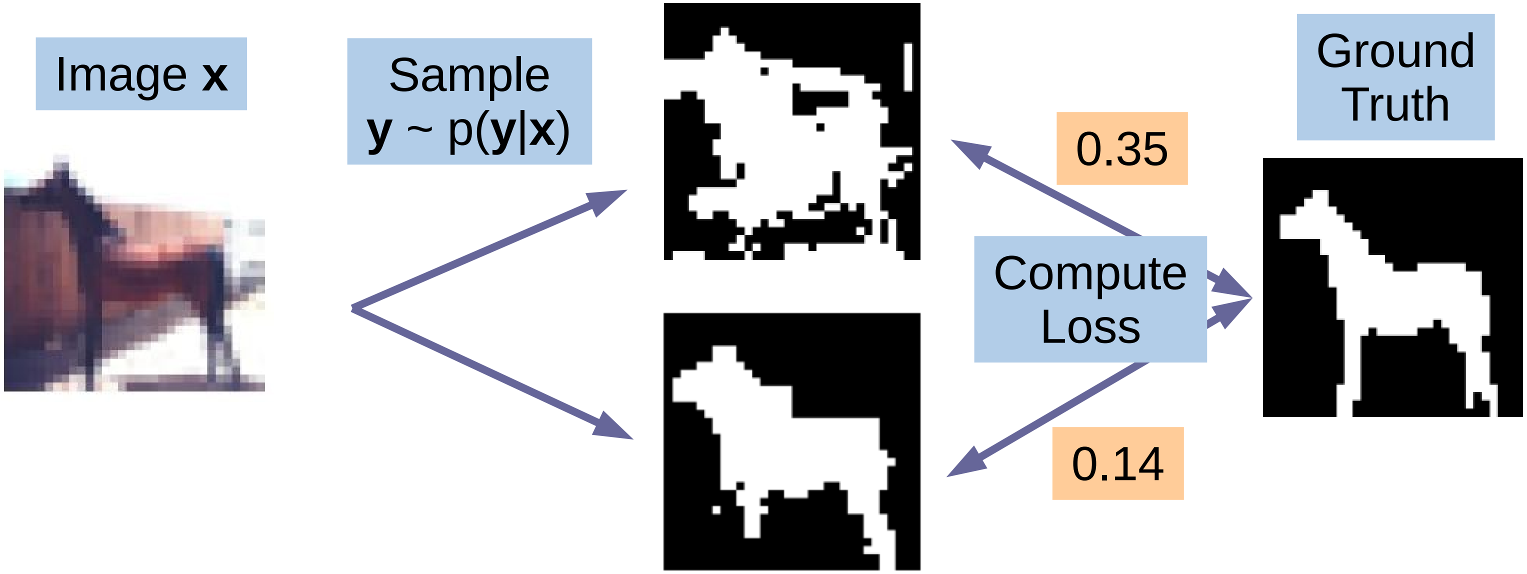


Learning by Minimizing Expected Loss

- Contrastive Divergence does not work well
- Expected loss objective

$$L = \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x};\theta)\ell(\mathbf{y},\mathbf{y}^*)$$

- Estimate gradient using a set of samples from $p(\mathbf{y}|\mathbf{x})$

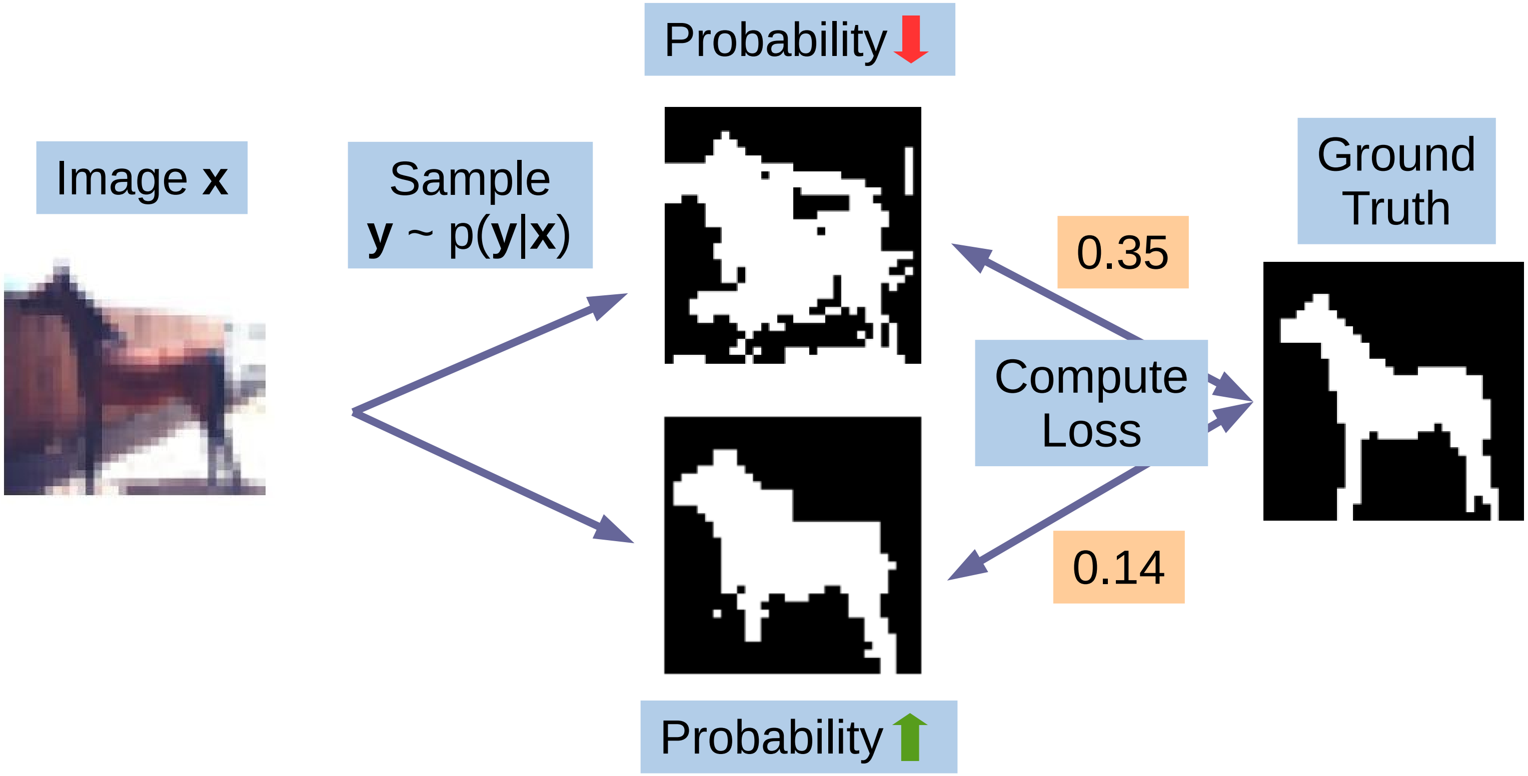


Learning by Minimizing Expected Loss

- Contrastive Divergence does not work well
- Expected loss objective

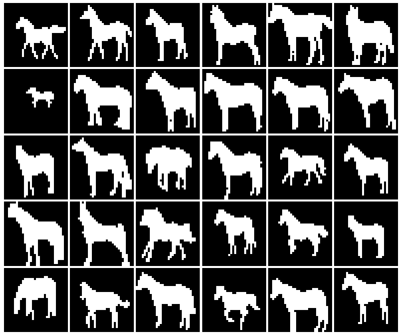
$$L = \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x};\theta)\ell(\mathbf{y},\mathbf{y}^*)$$

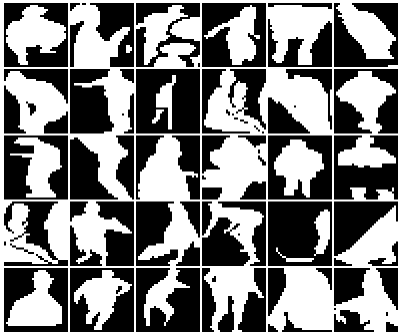
- Estimate gradient using a set of samples from $p(\mathbf{y}|\mathbf{x})$

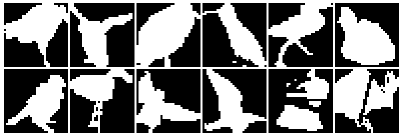


Datasets and Settings

- Weizmann horse dataset
- PASCAL VOC 2011: image inside the bounding box
  - Class “person” and class “bird”
- All images resized to 32x32
- *T*=1, Intersection Over Union (IOU) performance measure





Experiment I

- Train RBM independently (unsupervised)

Method	Horse IOU	Bird IOU	Person IOU
Unary Only	0.5119	0.5055	0.4979
iPW	0.5736	0.5585	0.5094
iPW+RBM	0.6722	0.5647	0.5126

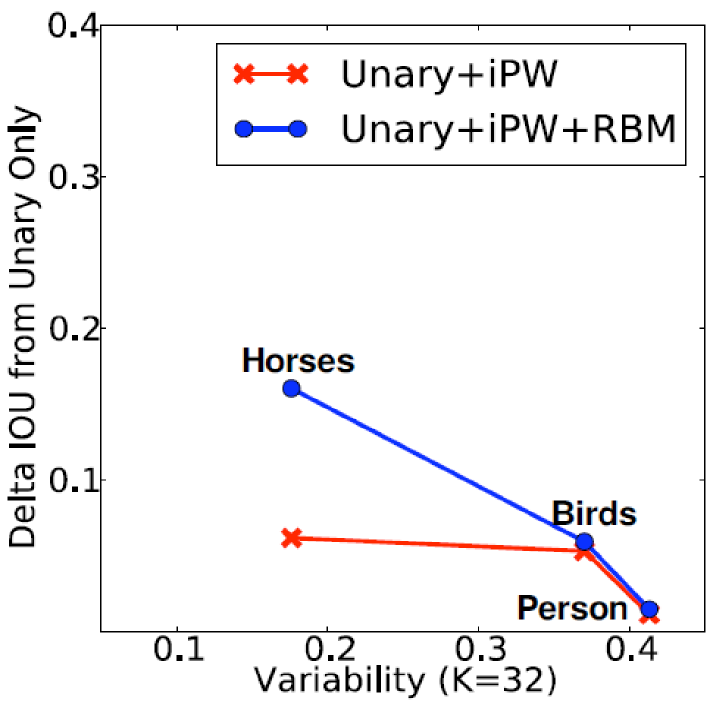
- Adding an RBM always helps
  - But not equally on different datasets

Experiment I Analysis: Dataset Variability

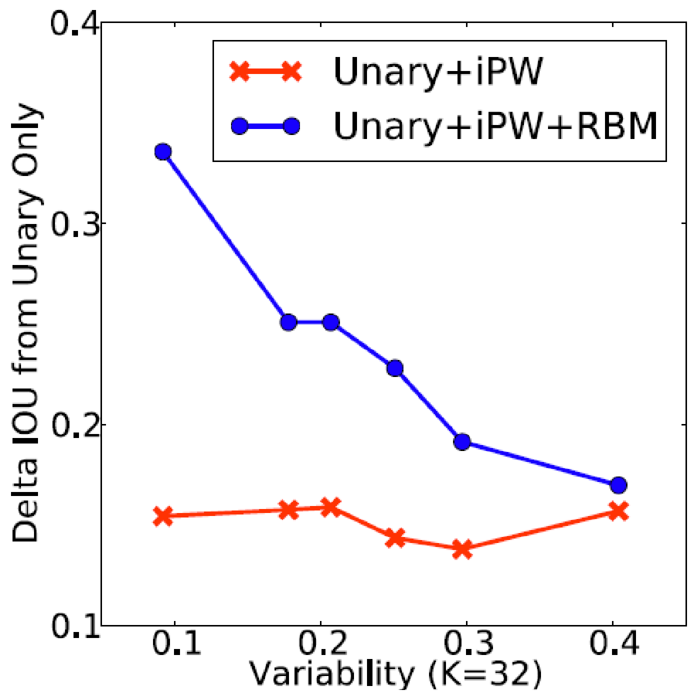
- Dataset variability measure

Clustering → Intra-cluster entropy → Weighted average

  - Person & Birds are harder than horses



Real Datasets



Synthetic Datasets

Experiment II and III

- Jointly learning RBM parameters by minimizing expected loss

Method	Horse IOU	Bird IOU	Person IOU
iPW+RBM	0.6722	0.5647	0.5126
iPW+jRBM	**0.6990**	**0.5773**	**0.5253**

Experiment II and III

- Jointly learning RBM parameters by minimizing expected loss

Method	Horse IOU	Bird IOU	Person IOU
iPW+RBM	0.6722	0.5647	0.5126
iPW+jRBM	**0.6990**	**0.5773**	**0.5253**

- Making the RBM hidden bias conditioned on the image

Method	Bird IOU	Person IOU
PW	0.5321	0.5082
iPW	0.5585	0.5094
iPW+jRBM	0.5773	**0.5253**
iPW+ijRBM	**0.5858**	**0.5252**

Examples


GT
U+P
CHO PP
Most Improvement
Average Improvement
Least Improvement

Conclusion and Future Work

- Theoretical contribution
  - Relationship between RBMs and Pattern Potentials
- Algorithmic contribution
  - Inference and learning algorithms for CHOPP-augmented CRFs
- Empirical contribution
  - Dataset variability measure

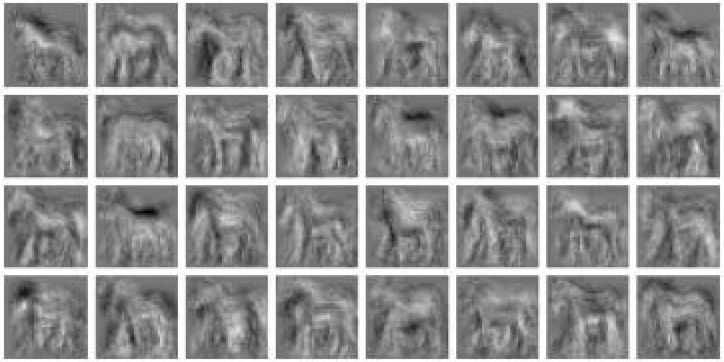
- Looking forward:
  - Convolutional and deeper models
  - Fully explore the variants of CHOPP
  - Challenge: lack of labeled data

Q & A
Exploring Compositional High Order Pattern Potentials for Structured Output Learning

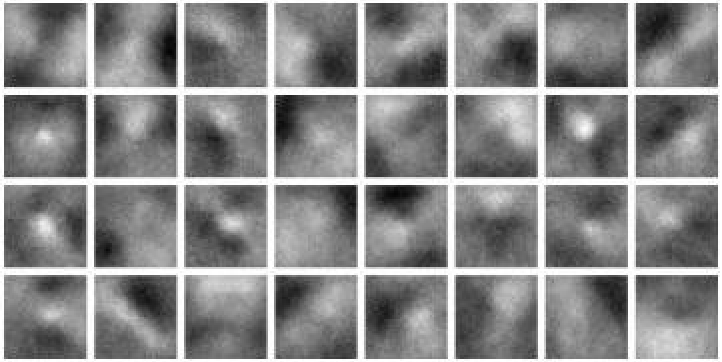
Yujia Li, Daniel Tarlow*, Richard Zemel

University of Toronto
*Now at Microsoft Research Cambridge
June 25, 2013

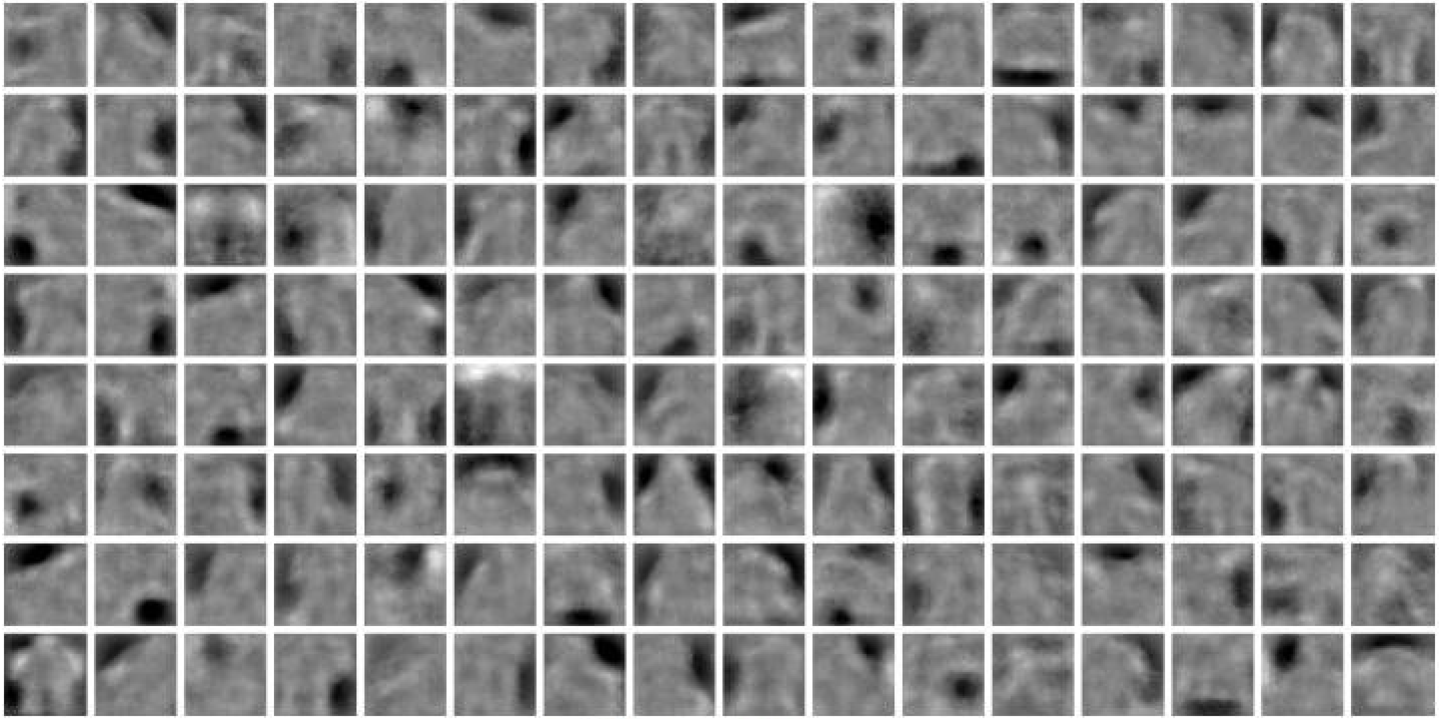
Learned Patterns



(a) Horse filters



(b) Bird filters



(c) Person filters.