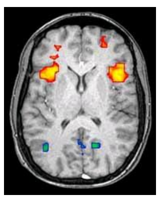
Deep Machine Learning: Panel Presentation

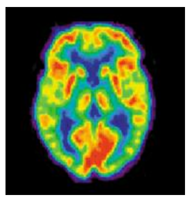
Honglak Lee
Computer Science & Engineering Division
University of Michigan

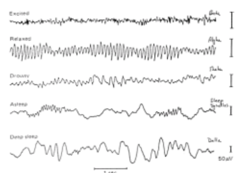
MSR Faculty Summit
7/16/2013

Mining meaningful structures from data

- **Multimedia (images, videos, speech, music, text, etc.)**
- **Healthcare data (medical imaging data, preoperative conditions, time series measurements, etc.)**


fMRI


PET scan


EEG

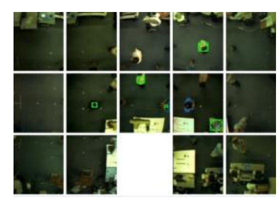

Ultra sound

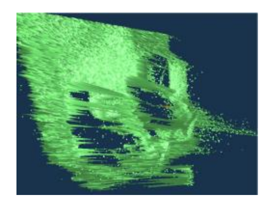
- **Multi modal sensor networks (e.g., robotics, surveillance, etc.)**


Visible light image


Audio


Thermal Infrared


Camera array


3d range scans

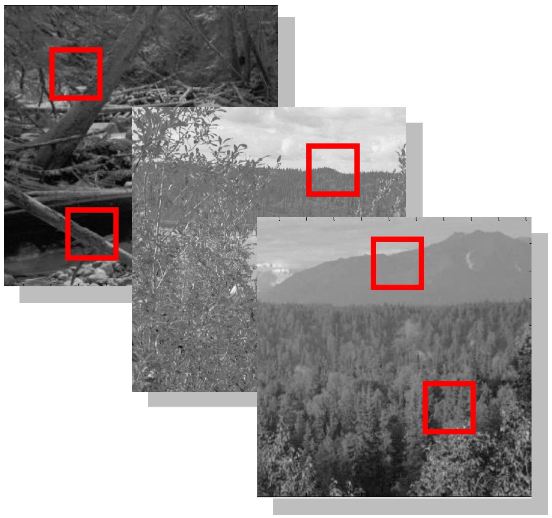
Learning Representations

- Key ideas:
  - **Unsupervised Learning**: Learn statistical structure or correlation of the data from unlabeled data (and some labeled data)
  - **Deep Learning**: Learn multiple levels of representation of increasing complexity/abstraction.
  - The learned representations can be used as features in **supervised** and **semi-supervised** settings.
- I will also talk about how to go beyond supervised (or semi-supervised) problems, such as:
  - **Weakly supervised learning**
  - **Structured output prediction**

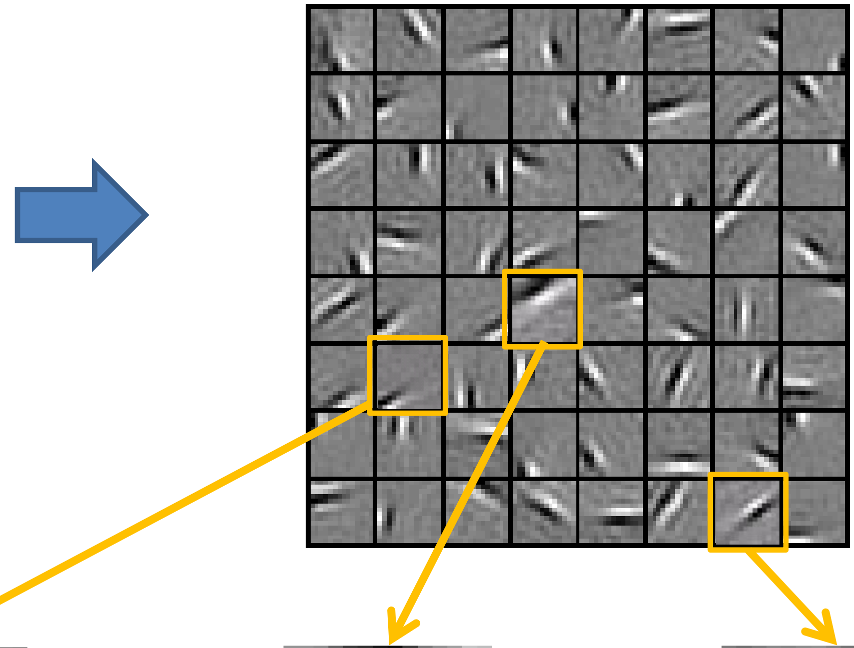
Unsupervised learning with sparsity

[NIPS 07; ICML 07; NIPS 08]

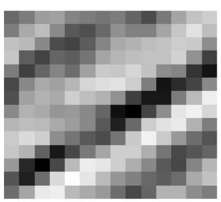
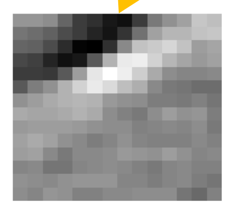
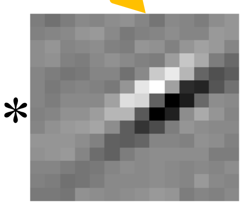
Natural Images



Learned bases: “Edges”



Test example

 $\sim 0.8 *$  $+ 0.3 *$  $+ 0.5 *$ 

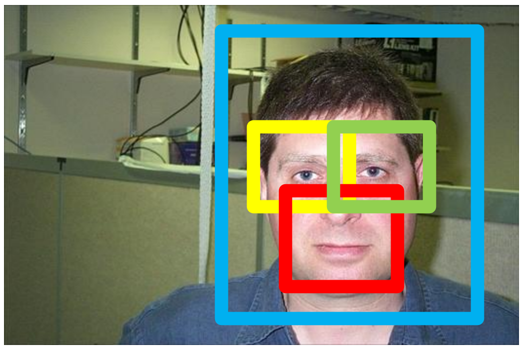
$$x \sim 0.8 * b_{36} + 0.3 * b_{42} + 0.5 * b_{65}$$

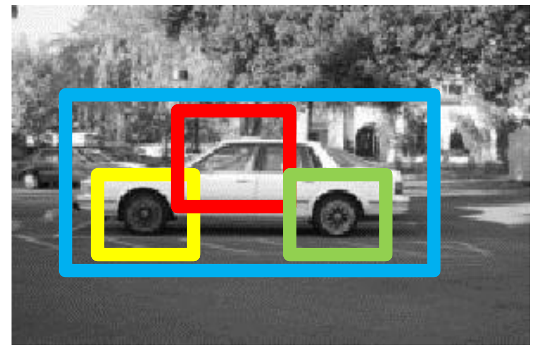
[0, 0, …, 0, **0.8**, 0, …, 0, **0.3**, 0, …, 0, **0.5**, …]
= coefficients (feature representation)

Compact & easily interpretable

Learning object representations

- Learning objects and parts in images

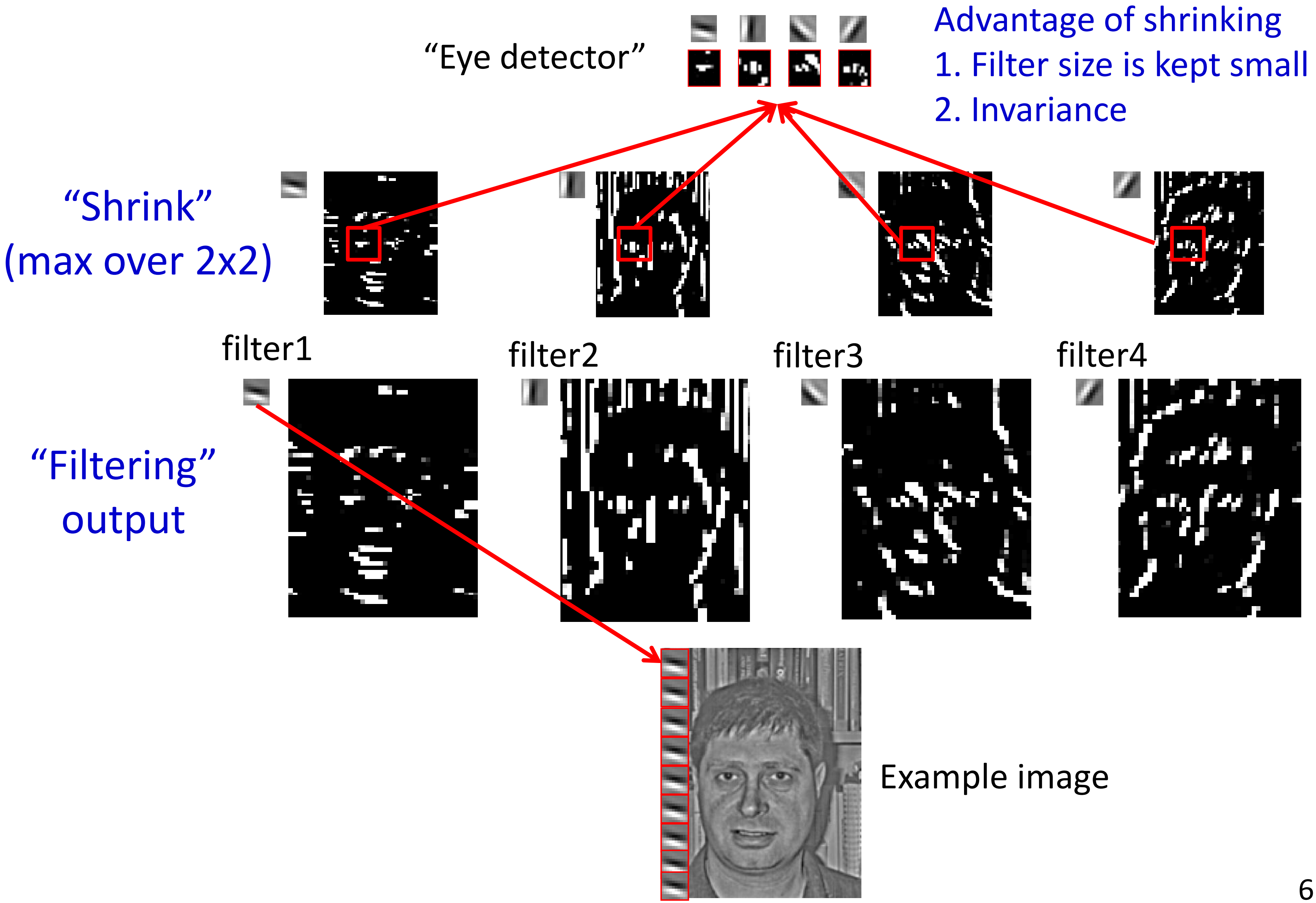




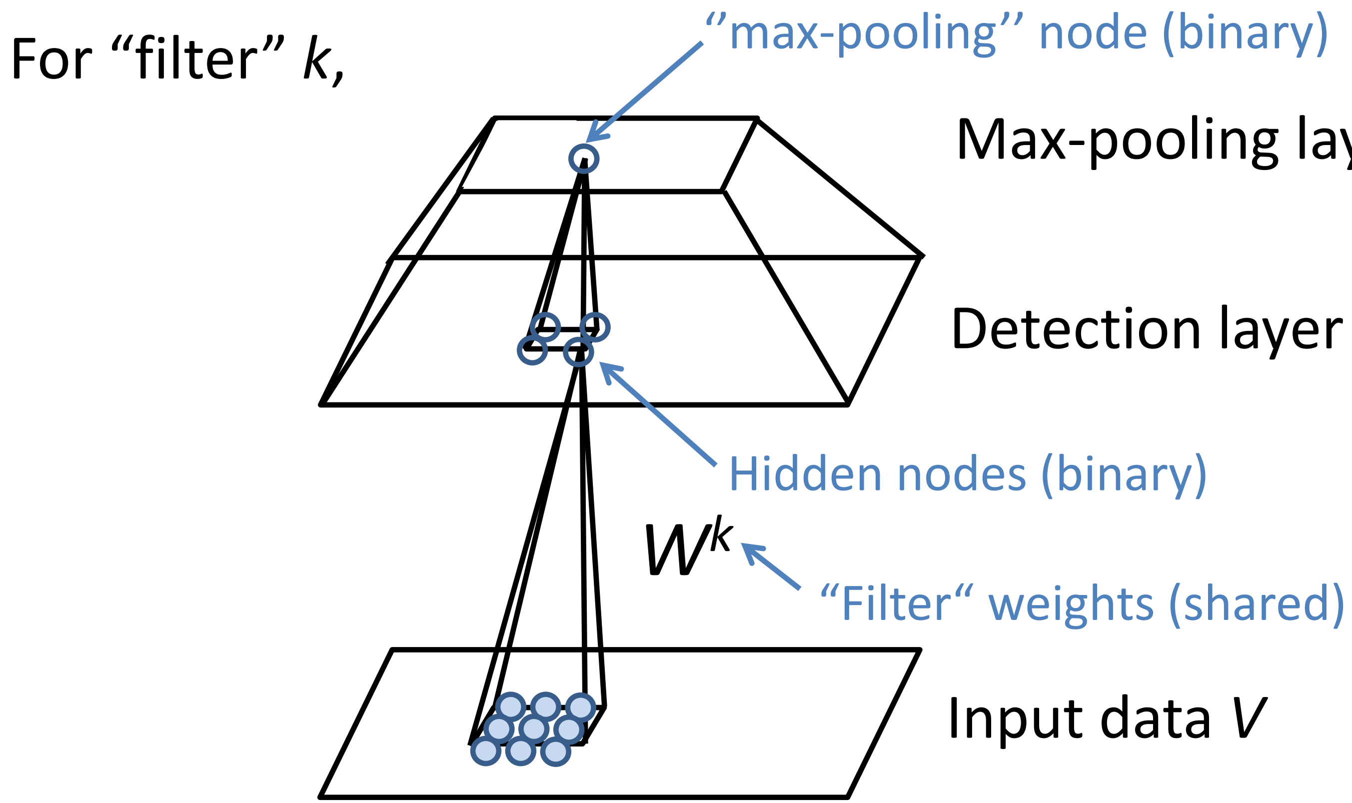
- Large image patches contain interesting higher-level structures.
  - E.g., object parts and full objects

- Challenge: high-dimensionality and spatial correlations

Illustration: Learning an “eye” detector



Convolutional RBM (CRBM) [ICML 2009]

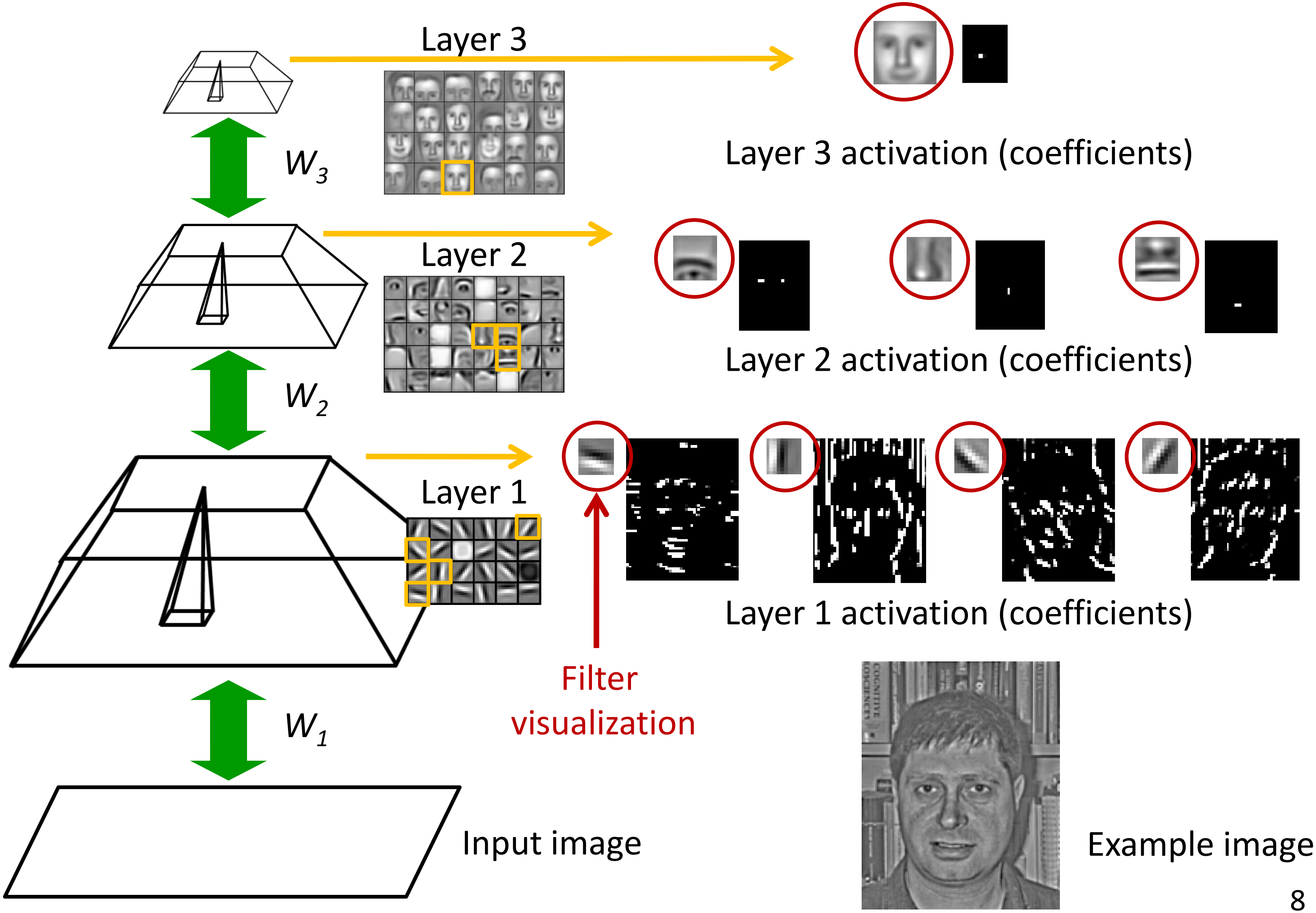


$$P(\mathbf{v}, \mathbf{h}) \propto \exp\left(\sum_{i,j,k} h_{i,j}^{k} (\tilde{W}^{k} * v)_{i,j}\right)$$

$$\text{subj. to} \quad \sum_{(i,j)\in\text{“}cell(y)\text{''}} h_{i,j}^{k} \leq 1, \forall k, y.$$

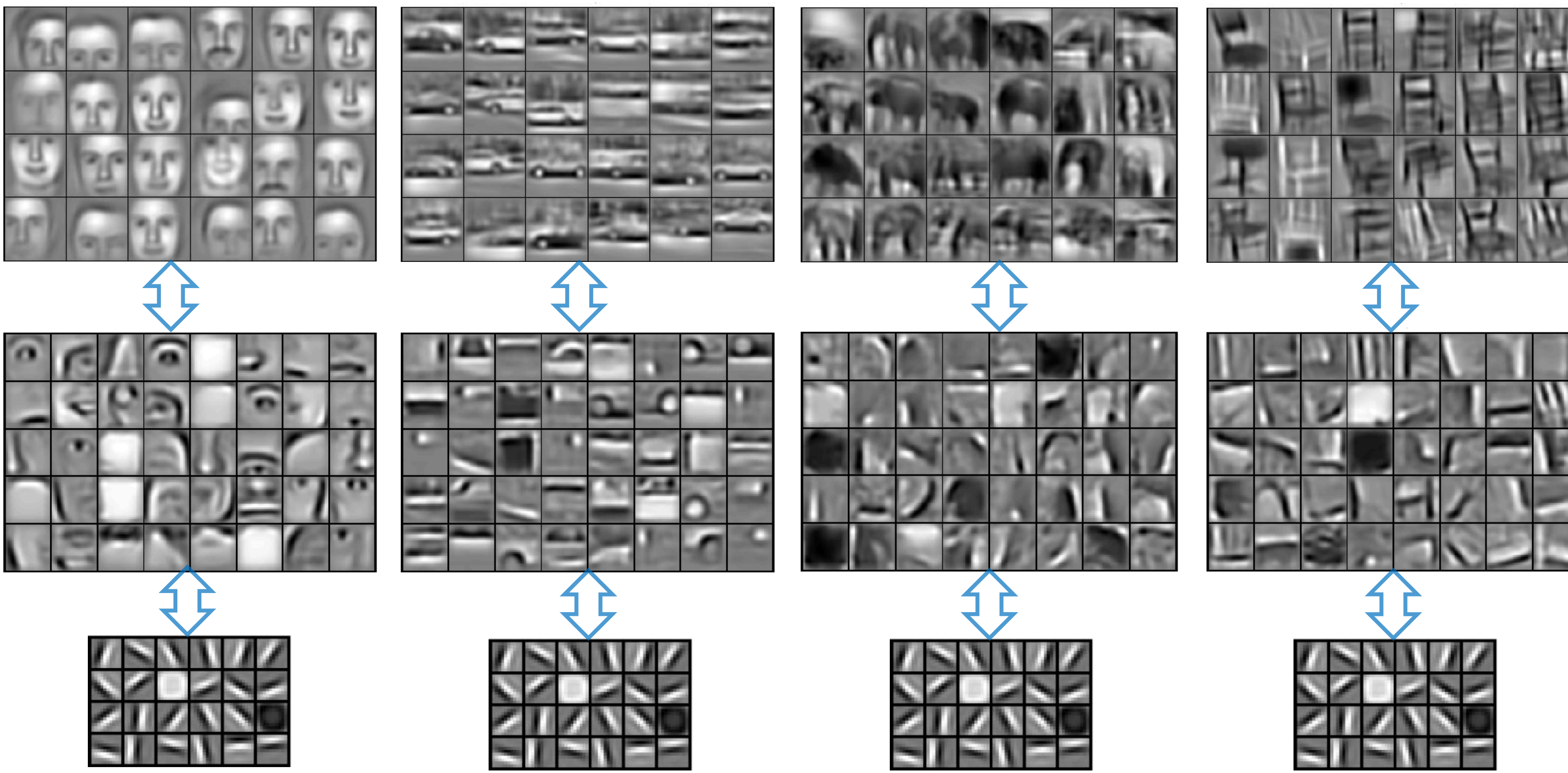
- RBM (probabilistic model)
- Convolutional structure
- Probabilistic max-pooling ("mutual exclusion")

Convolutional deep belief networks illustration



Learning object-part decomposition

Faces Cars Elephants Chairs



Applications

- Classification (ICML 2009, NIPS 2009, ICCV 2011, Comm. ACM 2011)
- Verification (CVPR 2012)
- Image alignment (NIPS 2012)
- The algorithm is applicable to other domains, such as audio (NIPS 2009)

Ongoing Work

- Investigating theoretical connections and efficient training (ICCV 2011)
- Robust feature learning with weak supervision (ICML 2013)
- Representation learning with structured outputs (CVPR 2013)
- Learning invariant representations (ICML 2009; NIPS 2009; ICML 2012)
- Multi-modal feature learning (ICML 2011)
- Life-long representation learning (AISTAST 2012)

Ongoing Work

- **Investigating theoretical connections and efficient training (ICCV 2011)**
- Robust feature learning with weak supervision (ICML 2013)
- Representation learning with structured outputs (CVPR 2013)
- Learning invariant representations (ICML 2009; NIPS 2009; ICML 2012)
- Multi-modal feature learning (ICML 2011)
- Life-long representation learning (AISTAST 2012)

Theoretical Connections and Efficient Training

- Connections between unsupervised learning methods
  - Clustering vs. distributed representation [Coates, Lee, Ng, AISTATS 2011]
  - Can we develop better learning algorithms using the links?
- Explore the connections between mixture models and RBMs.

GMM = Softmax Gaussian RBM < Activation-constrained RBM ≈ Sparse RBM < Convolutional sparse RBM

  - We provide an **efficient training method for RBMs via the connection.**
  - This is the first work showing that RBMs can be trained so that they are no worse than Gaussian Mixture models (GMMs).
- State-of-the-art results on object classification tasks.

Spherical Gaussian Mixtures is equivalent to RBM with softmax constraints

GMM = Softmax Gaussian RBM

$$P(\mathbf{v},\mathbf{h}) = \frac{1}{Z}\exp(-E(\mathbf{v},\mathbf{h}))$$

$$E(\mathbf{v},\mathbf{h}) = \frac{1}{2\sigma^2}\sum_i (v_i - c_i)^2 - \frac{1}{\sigma}\Big(\sum_{i,j} v_i W_{ij} h_j + \sum_j b_j h_j\Big)$$

$$\text{subj. to} \quad \sum_j h_j \leq 1$$

Gaussian Softmax RBM
= GMM with shared covariance $\sigma^2\mathbf{I}$

Relaxing the constraints

GMM = Softmax Gaussian RBM < Activation-constrained RBM

$$P(\mathbf{v},\mathbf{h}) = \frac{1}{Z}\exp(-E(\mathbf{v},\mathbf{h}))$$

$$E(\mathbf{v},\mathbf{h}) = \frac{1}{2\sigma^2}\sum_i (v_i - c_i)^2 - \frac{1}{\sigma}(\sum_{i,j} v_i W_{ij} h_j + \sum_j b_j h_j)$$

$\cancel{\text{subj. to} \quad \sum_j h_j \leq 1}$ ~~Gaussian Softmax RBM~~

subj. to $\sum_{k=1}^{K} h_k \leq \boldsymbol{\alpha}$, *activation-constrained* RBM

Relaxing the constraints

GMM = Softmax Gaussian RBM < Activation-constrained RBM ≈ Sparse RBM

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} \exp(-E(\mathbf{v}, \mathbf{h}))$$

$$E(\mathbf{v}, \mathbf{h}) = \frac{1}{2\sigma^2} \sum_i (v_i - c_i)^2 - \frac{1}{\sigma} (\sum_{i,j} v_i W_{ij} h_j + \sum_j b_j h_j)$$

~~subj. to $\sum_j h_j \leq 1$~~ ~~Gaussian Softmax RBM~~

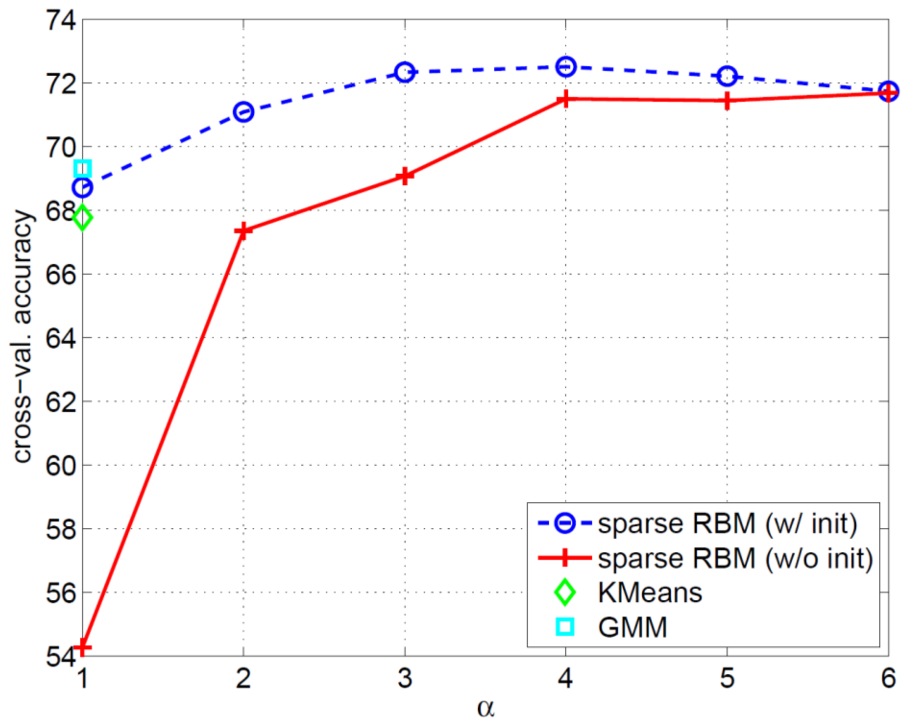
~~subj. to $\sum_{k=1}^{K} h_k \leq \boldsymbol{\alpha}$,~~ ~~*activation constrained* RBM~~

sparse RBM:
(regularize in training)
$\frac{1}{K} \sum_{k=1}^{K} h_k \approx \frac{\alpha}{K}$

Experiments – Analysis

[ICCV 2011]

- Effect of *sparsity* to the classification performance (Caltech 101).



  - The sparsity > 1/K showed the best CV accuracy.

- **Practical guarantee** that the sparse RBM lead to comparable or better classification performance than Gaussian mixtures.

Ongoing Work

- Investigating theoretical connections and efficient training (ICCV 2011)
- **Robust feature learning with weak supervision (ICML 2013)**
- Representation learning with structured outputs (CVPR 2013)
- Learning invariant representations (ICML 2009; NIPS 2009; ICML 2012)
- Multi-modal feature learning (ICML 2011)
- Life-long representation learning (AISTAST 2012)

Learning from scratch

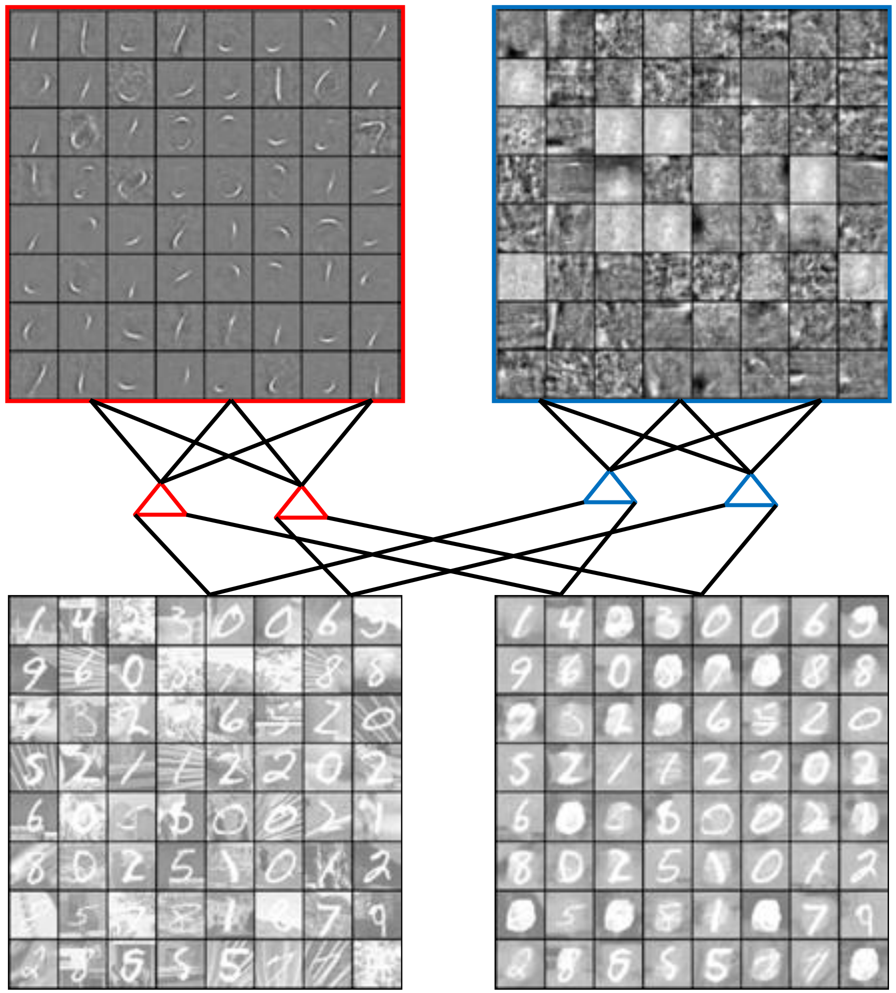
- Unsupervised feature learning
  - Powerful in **discovering** features from unlabeled data.
  - However, not all patterns (or data) are equally important.
    - When data contains lots of distracting factors, learning meaningful representations can be challenging.
- Feature selection
  - Powerful in **selecting** features from labeled data.
  - However, it assumes existence of discriminative features.
    - There may not be such features at hand.
- We develop a **joint model** for feature learning and feature selection
  - allows to learn **task-relevant high-level features** using (weak) supervision.

Experiments – visualizations

- Learning from noisy handwritten digits with PGBM

Learned task-relevant hidden unit weights: mostly ***pen-strokes***

Learned task-irrelevant hidden unit weights: noisy patterns

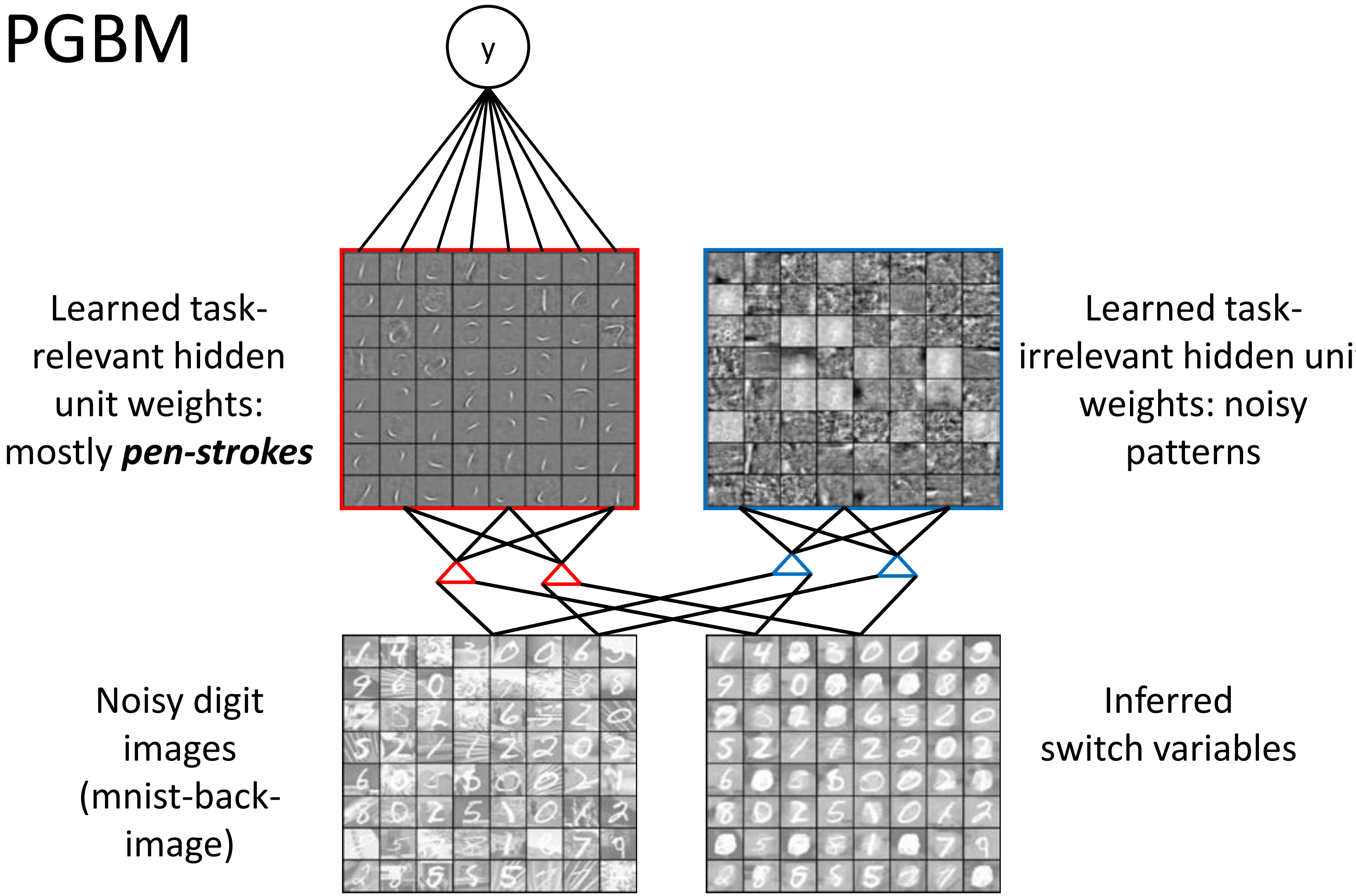


Noisy digit images (mnist-back-image)

Inferred switch variables

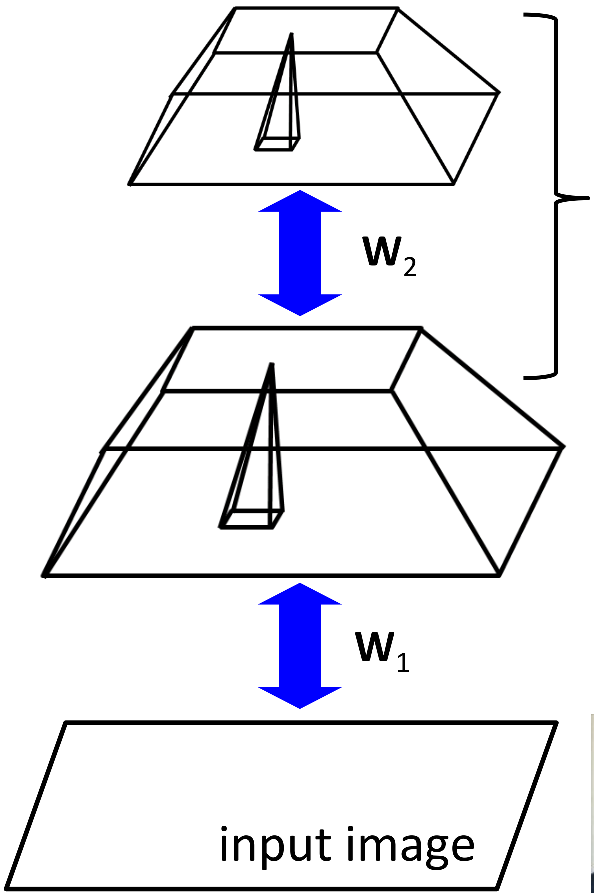
Experiments – visualizations

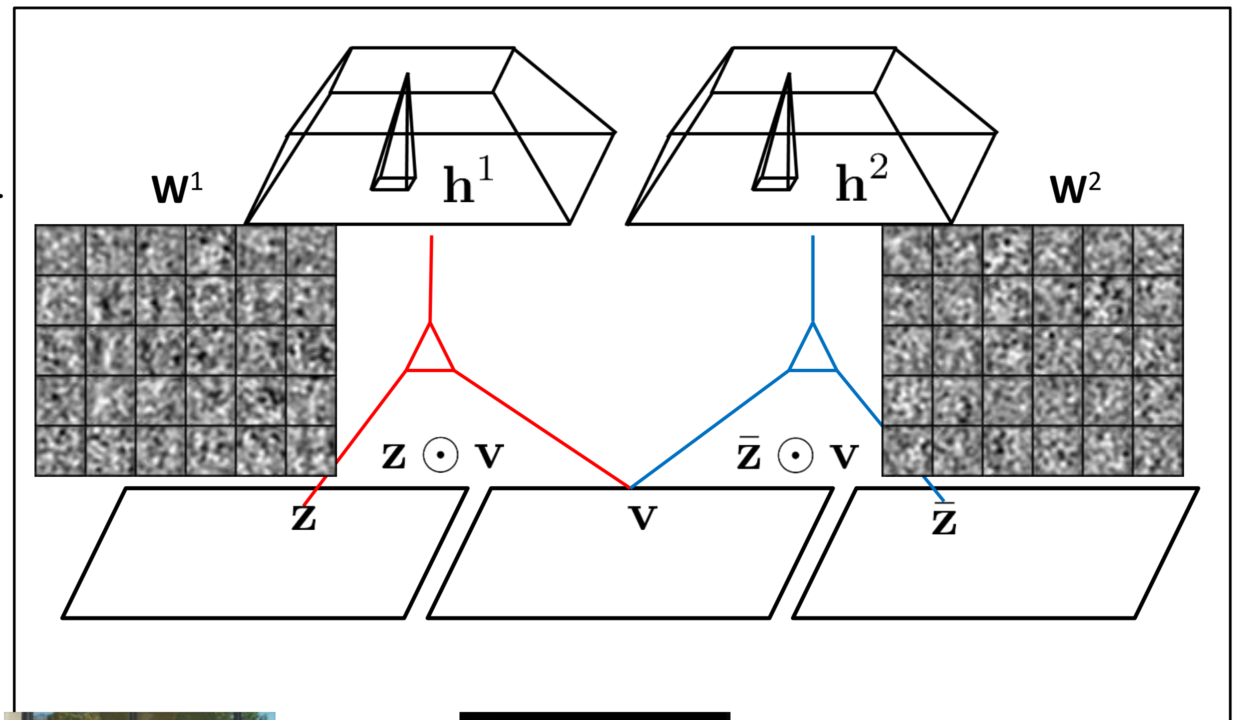
- Learning from noisy handwritten digits with PGBM

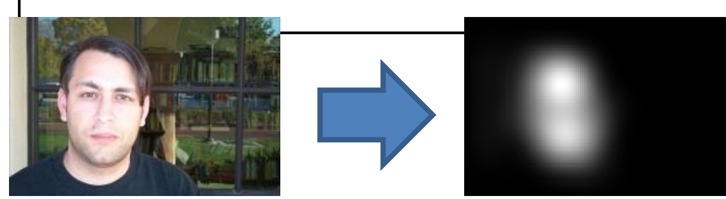


Convolutional Extensions

We can distinguish between task-relevant and irrelevant features with point-wise gating idea while feature learning.

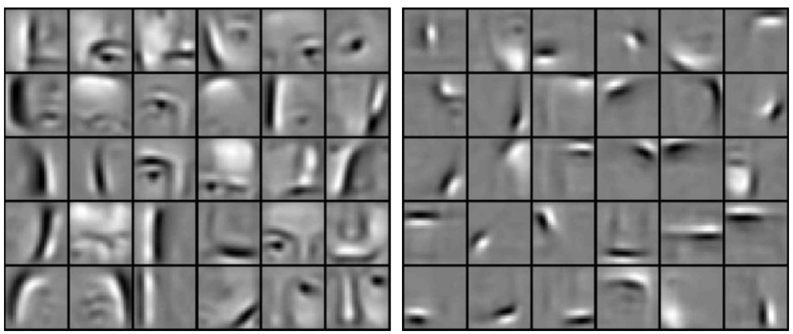




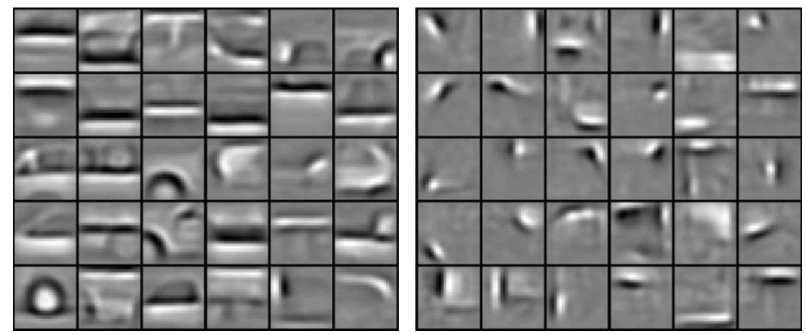


Experiments – weakly supervised object segmentation

- Learned set of filters (task-relevant/irrelevant)



Caltech101 - Faces

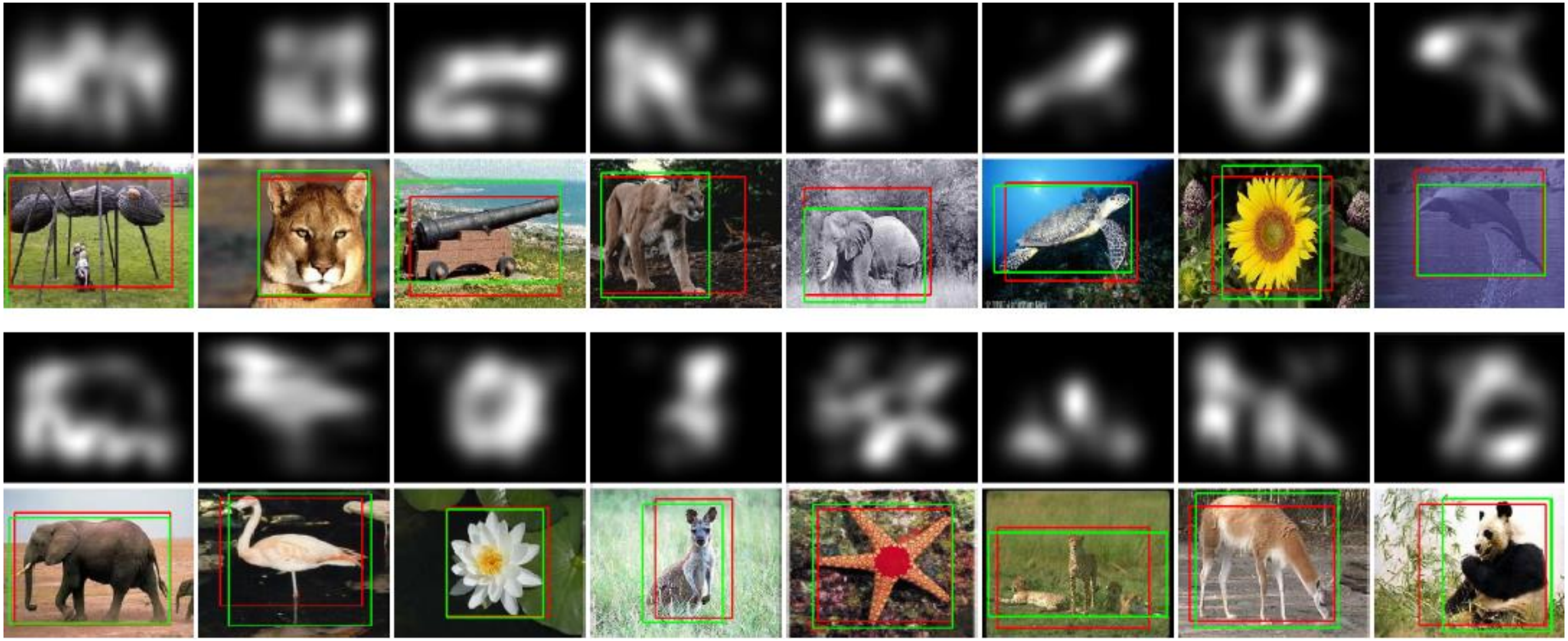


Caltech101 – car side

- (Weakly supervised) object localization

1st row: switch unit activation map,
2nd row: predicted and ground truth bounding box.

Experiments – weakly supervised object segmentation



1st row: switch unit activation map,
2nd row: predicted and ground truth bounding box.

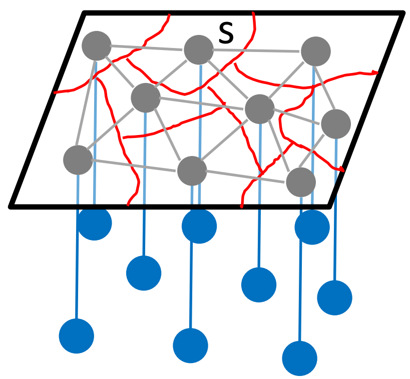
Ongoing Work

- Investigating theoretical connections and efficient training (ICCV 2011)
- Robust feature learning with weak supervision (ICML 2013)
- **Representation learning with structured outputs (CVPR 2013)**
- Learning invariant representations (ICML 2009; NIPS 2009; ICML 2012)
- Multi-modal feature learning (ICML 2011)
- Life-long representation learning (AISTAST 2012)

Enforcing Global and Local Consistencies for Structured Output Prediction

- Task: scene segmentation

Image | Over-segmentation | Target Output



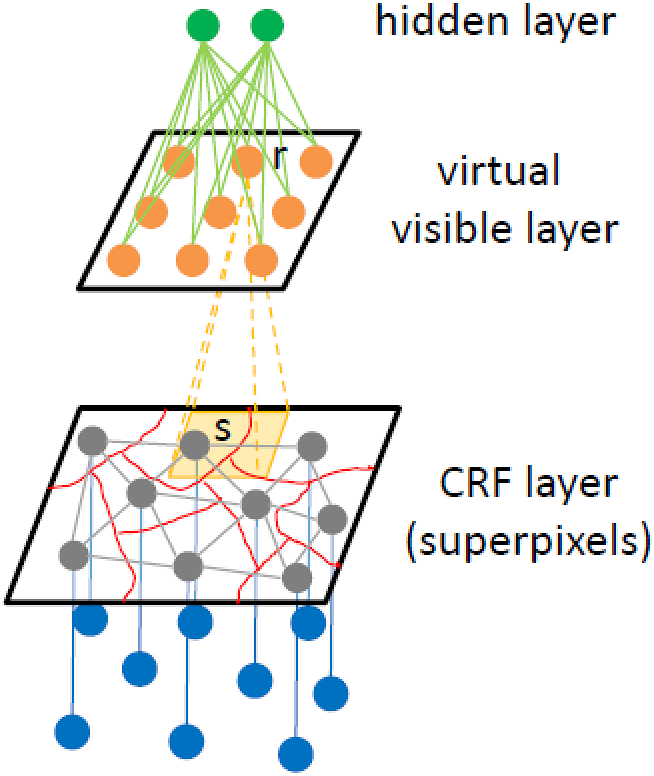
CRF with superpixels

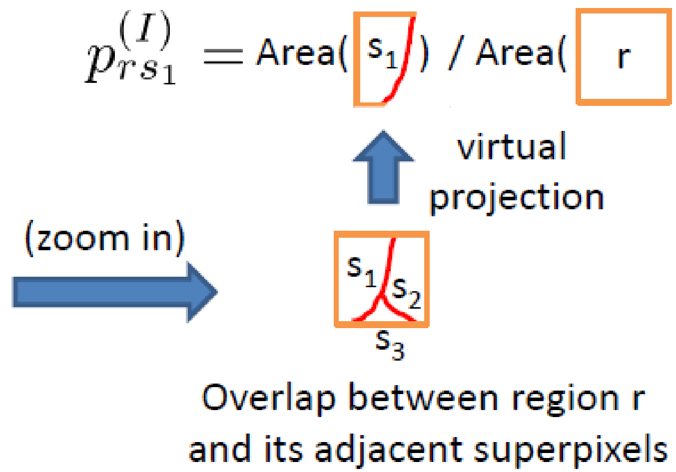
(CVPR 2013)

- Problem: only enforces local consistency
- Our model can enforce both local and global consistency

Combining Global and Local Consistencies for Structured Output Prediction

(CVPR 2013)





$$P(\mathbf{Y}|\mathbf{X}) = \frac{1}{Z}\sum_{\mathbf{h}} \exp\{-E(\mathbf{X}, \mathbf{Y}, \mathbf{h}; I)\}$$

$$E(\mathbf{X}, \mathbf{Y}, \mathbf{h}; I) = E_{\text{crf}}(\mathbf{X}, \mathbf{Y}) + E_{\text{rbm}}(\mathbf{Y}, \mathbf{h}).$$

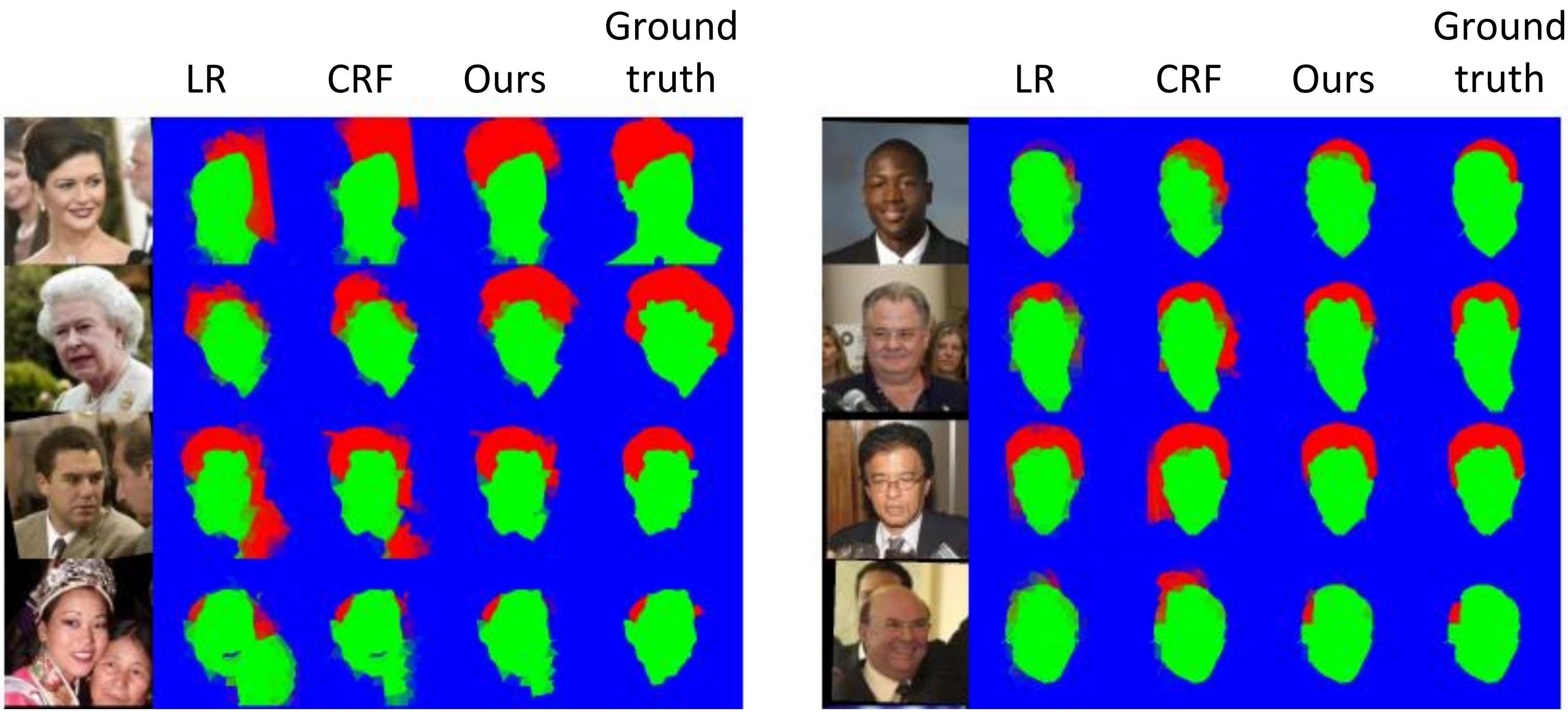
$$E_{\text{rbm}}(\mathbf{Y}, \mathbf{h}; I) = -\sum_{r=1}^{R^2}\sum_{l=1}^{L}\sum_{k=1}^{K} \bar{y}_{rl} W_{rlk} h_k - \sum_{k=1}^{K} b_k h_k - \sum_{r=1}^{R^2}\sum_{l=1}^{L} c_{rl}\bar{y}_{rl}$$

where $\bar{y}_{rl} \triangleq \sum_{s=1}^{S(I)} p^{(I)}_{rs} y_{sl}$

Experimental results

- Visualization of segmentation

(CVPR 2013)



  - LR: singleton potential
  - CRF: singleton + pairwise potential
  - Ours: singleton + pairwise + RBM potential

Summary

- Generative learning of convolutional feature hierarchy
- Better training algorithms
- Learning representations with weak supervision
- Learning representations with structured outputs

Funding support:

- NSF, ONR, Google, Toyota, Bosch Research