Scalable Learning for Restricted Boltzmann Machines

Elnaz Barshan
Paul Fieguth

System Design Engineering
University of Waterloo
Waterloo, Canada

29 October 2014

Vision and Image Processing Group
Systems Design Engineering

Outline

- Motivation
- Related Work
- Proposed Approach
- Experimental Results
- Conclusion

Object Recognition

- Why Challenging?
  - Large intra-class variation

Image Models and Unlabeled Data

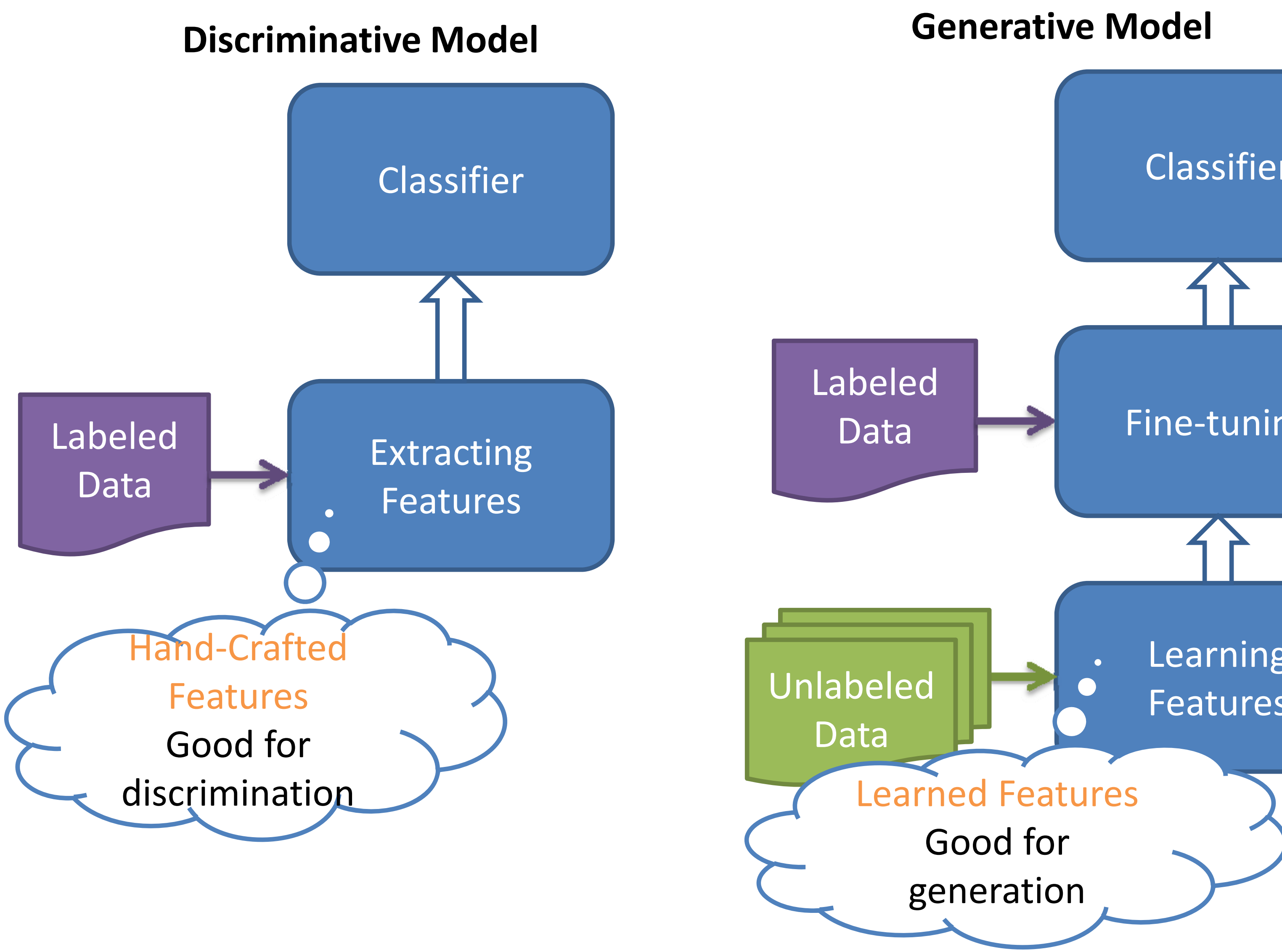
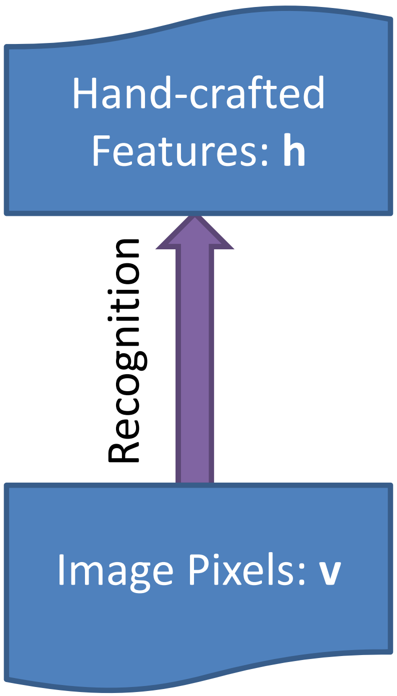
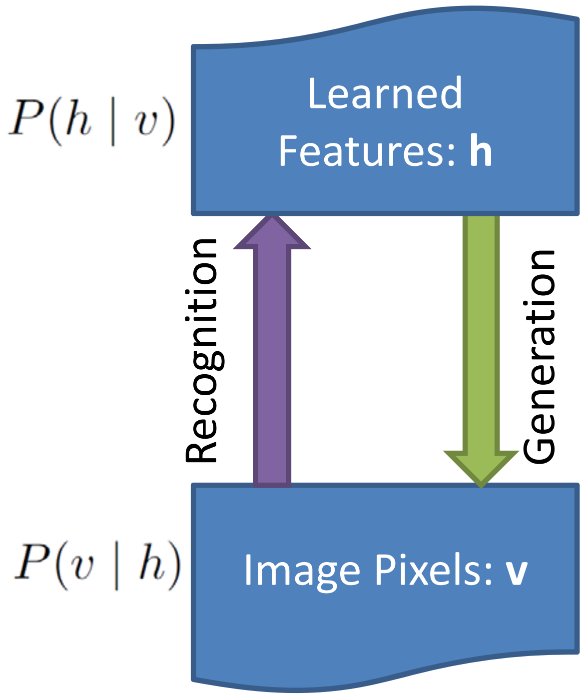


Image Models and Occlusion

Discriminative Model



Generative Model



$$P(v_{missing} \mid v_{observed}, h)$$

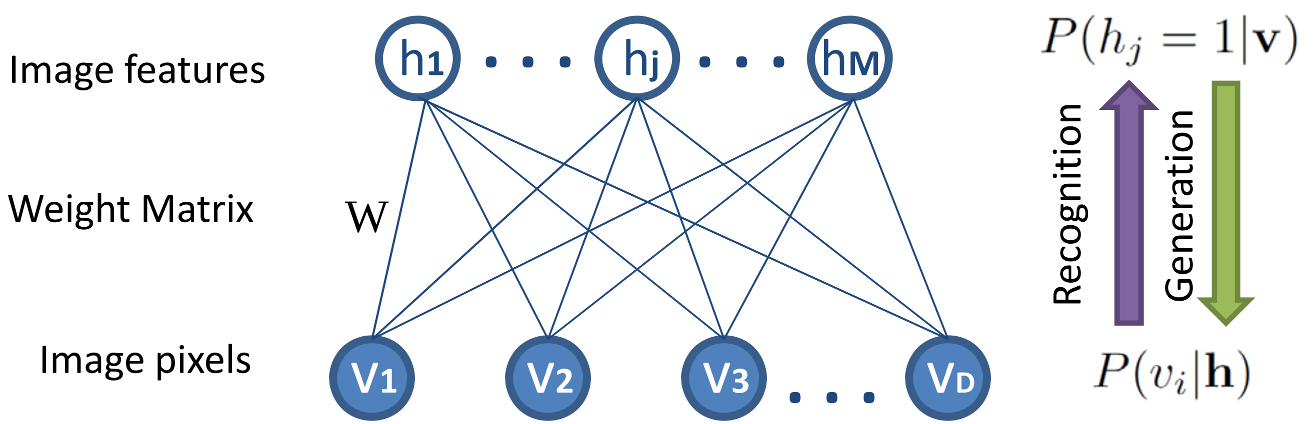
[Ranzato et al. 2011]

Outline

- Motivation
- Related Work
- Proposed Approach
- Experimental Results
- Conclusion

Generative Models in Visual Recognition

- Restricted Boltzmann Machines (RBMs)



- Learning:

$$W_{opt} = \underset{W}{\operatorname{argmax}}\, P(\mathbf{v})$$

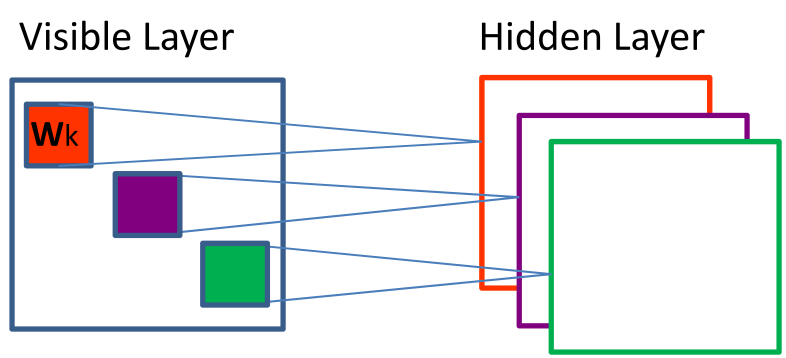
[Hinton 2002]

Basic RBM Issues

- Large Number of Parameters
  - Grows roughly quadratically with the image size
  - Expensive weight learning procedure
  - A threat to good generalization of the model
- Solution:
  - Using a weight sharing scheme

✓ Convolutional Architectures

✓ Translation invariance is frequently violated.



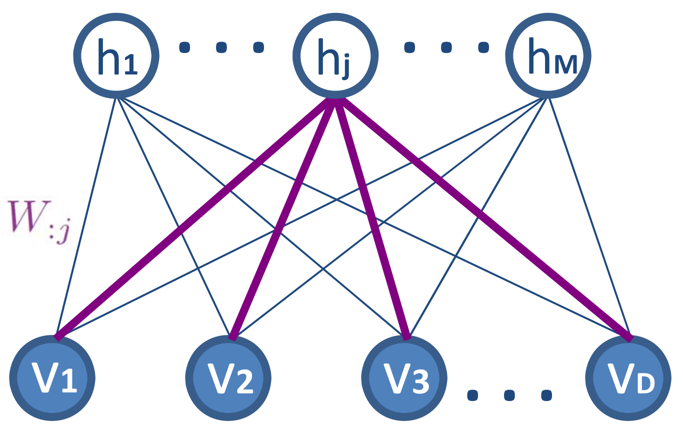
[Lee et al. 2009]

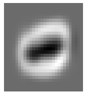
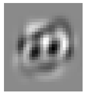
Outline

- Motivation
- Related Work
- Proposed Approach
- Experimental Results
- Conclusion

Proposed Strategy I

- Defining network weights as linear combinations of a set of predefined filters.



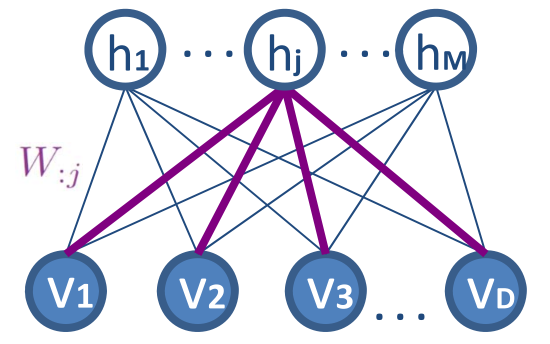
$W_{:j} = \alpha_{1j}$  $+ \cdots + \alpha_{Pj}$ 

✓The number of parameters becomes independent of the size of the image.

Proposed Strategy II

- Given a filter bank $F = \{\mathbf{f}^1, \mathbf{f}^2, ..., \mathbf{f}^P\}$:

$$W_{:j} := \sum_{k=1}^{P} \alpha_{kj} \mathbf{f}^k \qquad where\ P \ll D$$



- Learning the weights α_{kj} of the filters:

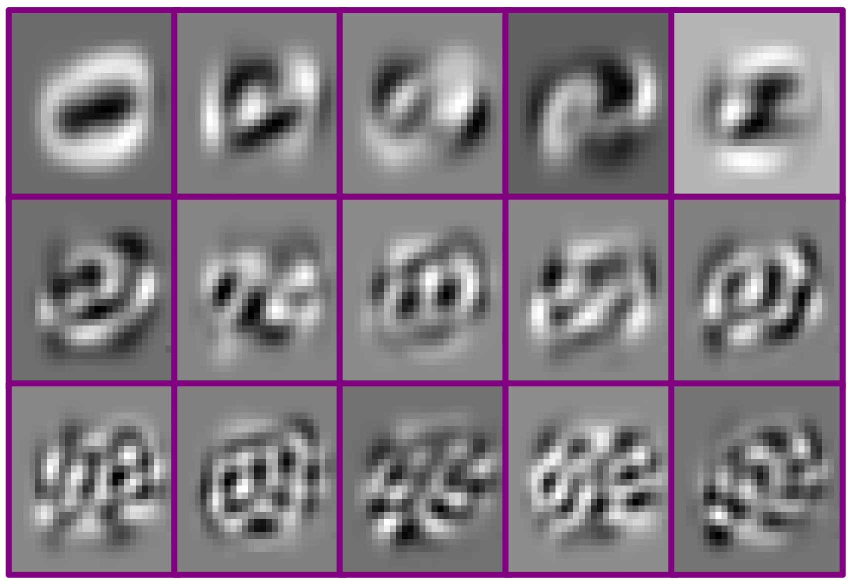
$$\frac{\partial log P(\mathbf{v})}{\partial \alpha_{kj}} = \sum_{i \in vis} \frac{\partial log P(\mathbf{v})}{\partial w_{ij}} \frac{\partial w_{ij}}{\partial \alpha_{kj}}$$

$$= \sum_{i \in vis} (< v_i h_j >_{data} - < v_i h_j >_{model}) f_i^k$$

Eigen-RBM

- Eigen-RBM
  - Filter Bank: Top eigenvectors of the covariance matrix of the training data.

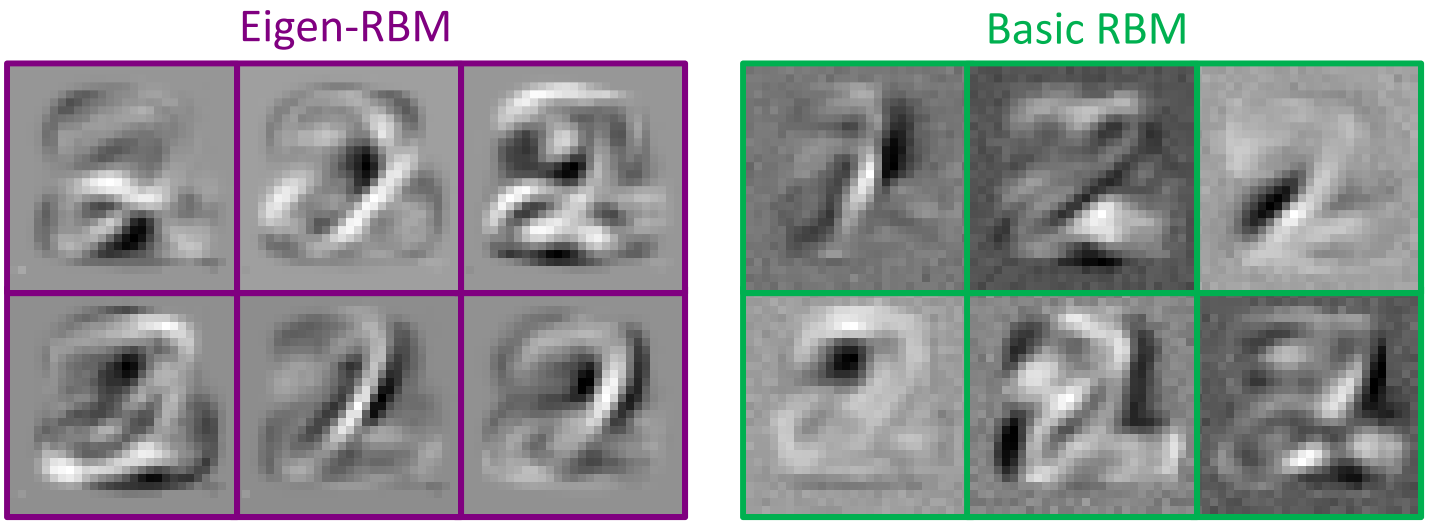
Eigendigits Filter bank



✓ Eigendigits capture information at a variety of scales from coarse to fine.

Eigen-RBM and Basic RBM

- The learned filters for a single digit class training data



✓ Although noise-reduction is not the objective, the learned Eigen-RBM filters are less noisy.

Outline

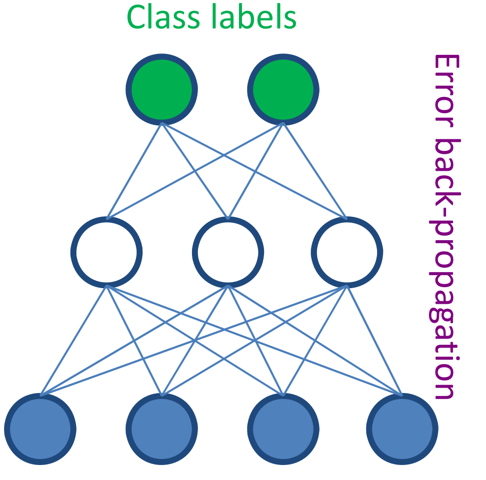
- Motivation
- Related Work
- Proposed Approach
- Experimental Results
- Conclusion

Setup

- Dataset
  - Small MNIST: 6000 training, 1000 test
  - 28x28 images of hand-written digits

- Classifier
  - 1NN

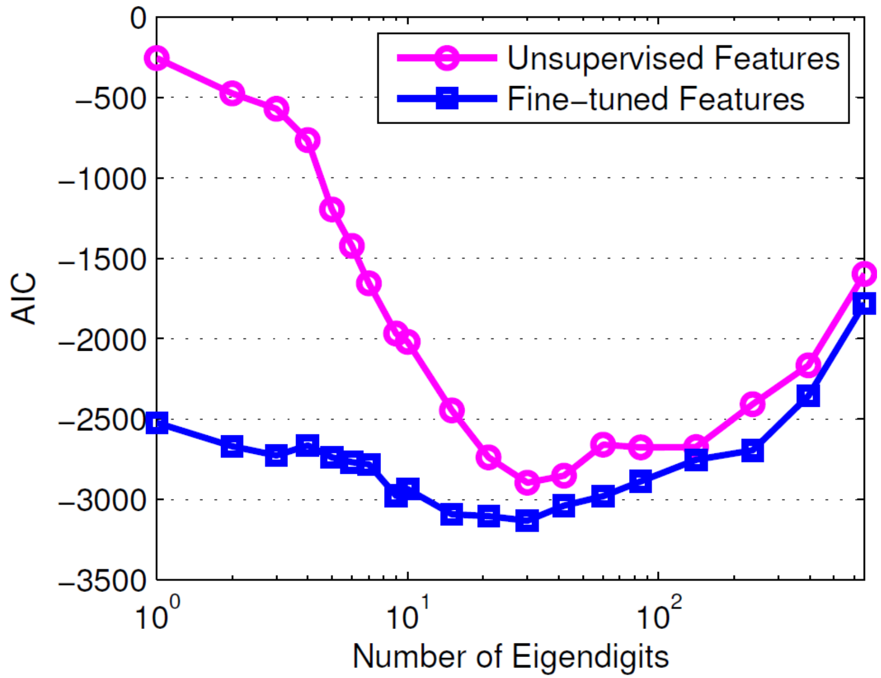
- Feature learning
  - Unsupervised
  - Fine-tuned using error back-propagation



Eigen-RBM Analysis

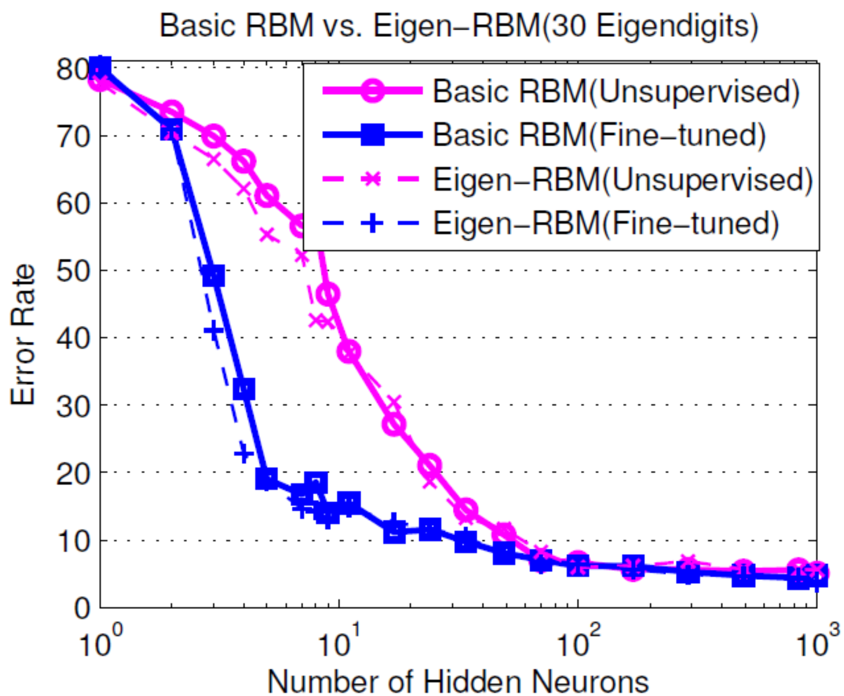
- Determining the optimum number of eigendigits
  - Akaike Information Criterion (AIC)

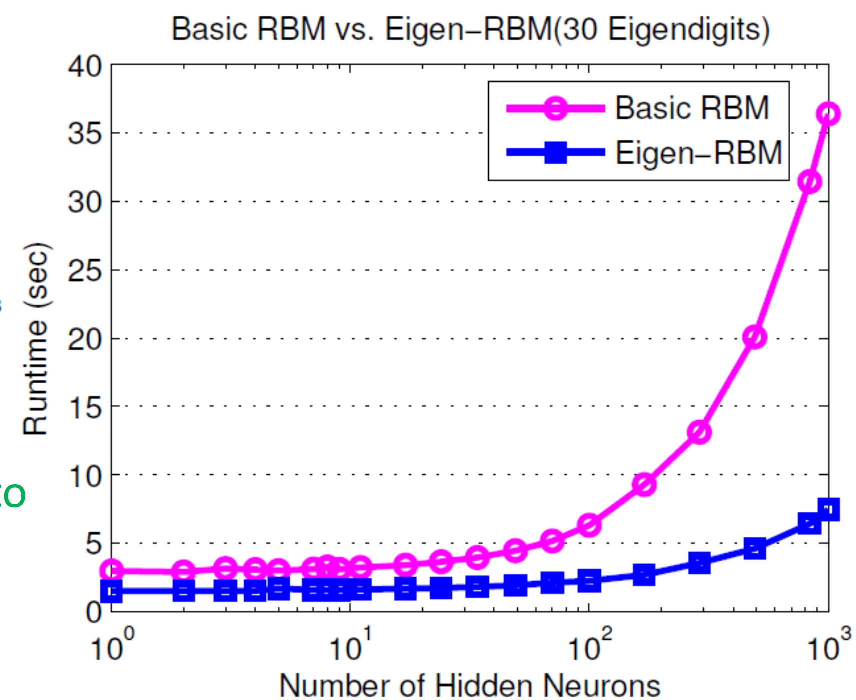
$$AIC = 2P + Nln(Error_{avrg})$$



✓ We use the top 30 Eigendigits of the data

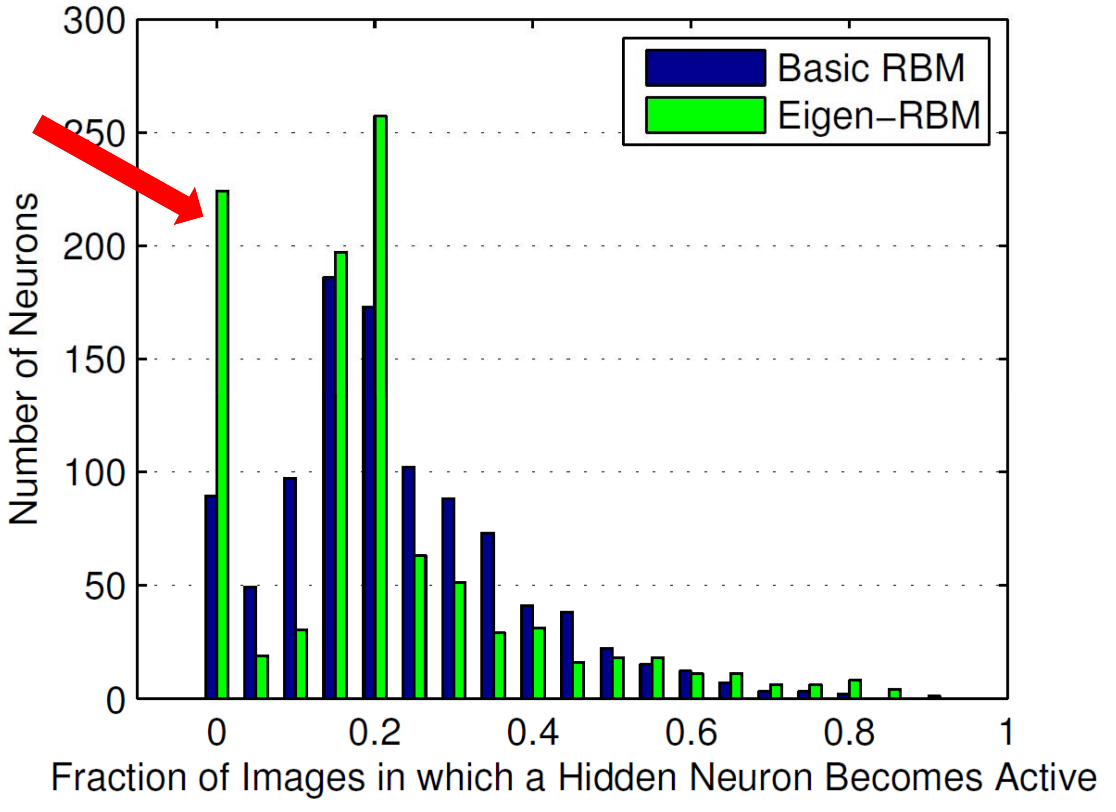
Recognition Performance





- ✓ Eigen-RBM has a similar performance to basic RBM, but with much less training time.

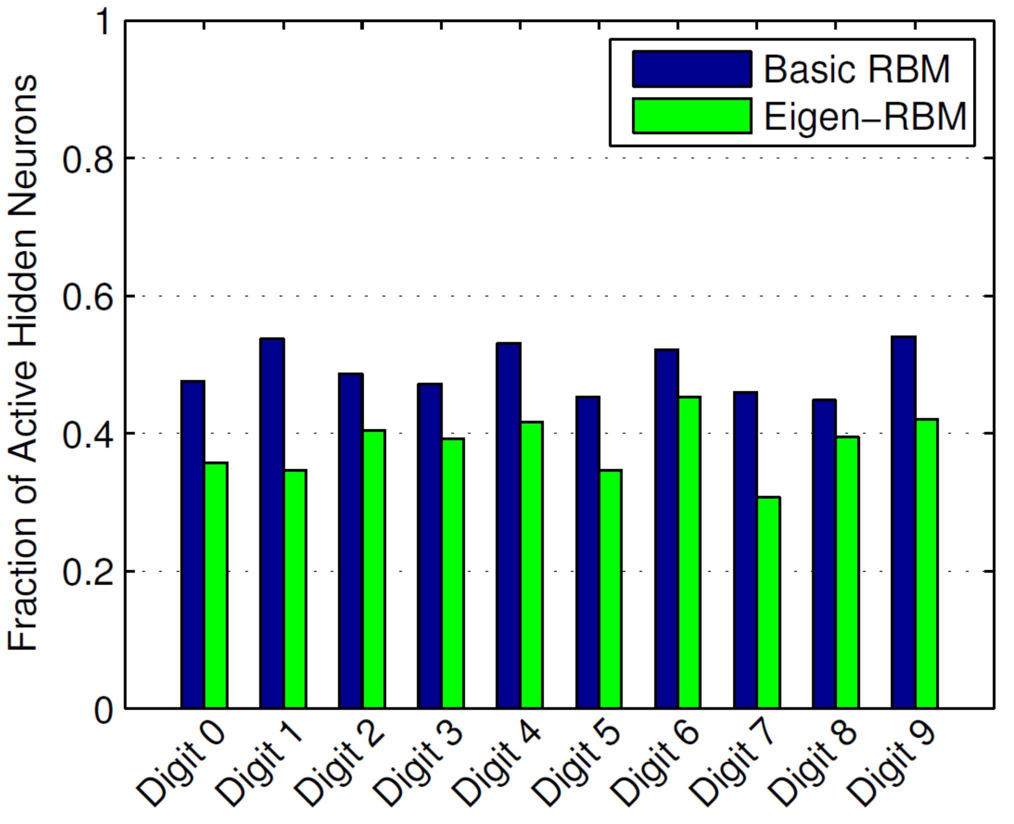
Sparsity in Recognition



✓ Eigen-RBM representation is sparser than that of basic RBM.

Sample Generation

Sparsity in Sample Generation



✓ Eigen-RBM generates similar or better samples with more sparse representations than basic RBM.

Outline

- Motivation
- Related Work
- Proposed Approach
- Preliminary Results
- Conclusion

Conclusion

- Eigen-RBM:
  - Scalable weight learning algorithm
  - Number of parameters is independent of the image size
- Compared to Basic RBM:
  - Similar or better performance
  - Much less training time
  - More sparse representations

Questions and Suggestions

Thank you!

Essentially, all Models are wrong, but some are useful.

George P. E. Box