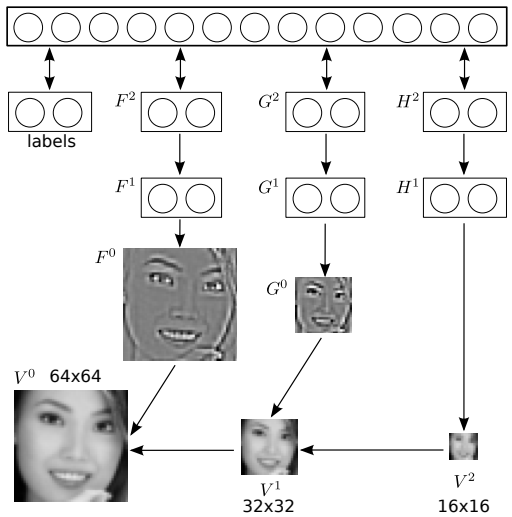
Multiresolution Deep Belief Networks

Charlie Tang
Abdel-rahman Mohamed
Department of Computer Science, University of Toronto, Canada

Introduction

- Coarse and fine resolutions of an image reveal different visual structures
- Multiresolution DBN combines a Laplacian Pyramid with deep learning
- Learning features from multiple resolutions and frequencies helps generalization
- Learning coarse structure from low resolution images leads to a better generative model

MrDBN



Algorithm 1 Inference and Generation for MrDBN

Inference:
Given an input image $\mathbf{v}^0$ of dimensionality $M \times M$, where $M = M_0 \times 2^K$, K an integer ≥ 2.

1: Low-pass filter $\mathbf{v}^0$ with a Gaussian kernel and downsample by a factor of 2 and 4 using bicubic interpolation, producing $\mathbf{v}^1$ and $\mathbf{v}^2$, respectively.
2: Upsample $\mathbf{v}^1$ and $\mathbf{v}^2$ by a factor of 2 using bicubic interpolation.
3: Subtract the upsampled images from the original to give the difference images: $\mathbf{f}^0$ and $\mathbf{g}^0$.
4: As in a standard DBN, use the RBM weights to compute approximate posterior distribution of a higher layer given the layer below. E.g. $p(\mathbf{f}^1|\mathbf{f}^0) = sigmoid\ (\mathbf{f}^{0\mathrm{T}}\Lambda^{\frac{1}{2}}\mathbf{W} + \mathbf{c})$.

Generation:
5: Run n iterations of block Gibbs sampling on the top level RBM.
6: Project down directed connections by computing the conditional distribution $p(\mathbf{x}^l|\mathbf{x}^{l+1})$, where x represents any of the three streams' hidden layer nodes.
7: Sample $\mathbf{x}^l \sim p(\mathbf{x}^l|\mathbf{x}^{l+1})$
8: Repeat step 6-7 until after the sampling of the layer of F^0, G^0, and V^2.
9: Reconstruct V^0:
$V^0 = UPSAMPLE(UPSAMPLE(\mathbf{v}^2) + \mathbf{g}^0) + \mathbf{f}^0$

Discriminative Experiments

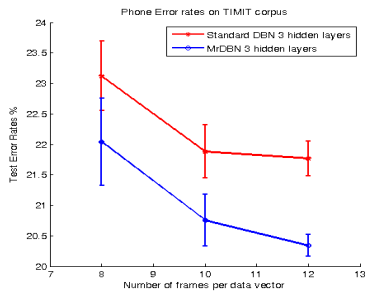
NORB Classification

Deep Belief Net [4]	8.3%
DBN (learning Λ) (this work)	7.4%
Deep Boltzmann Machine [2]	7.2%
MrDBN 32-16 (this work)	5.8%
3rd Order DBN [26]	5.2%
Tiled Convolutional Nets [27]	3.9%

	32x32	16x16	8x8	Combined
MrDBN [32-16-8]	11.2%	10.7%	9.2%	7.1%
MrDBN [32-16]	11.2%	6.9%	N/A	5.8%
DBN 32	7.4 %	N/A	N/A	N/A

- Recognition is improved by combining resolutions

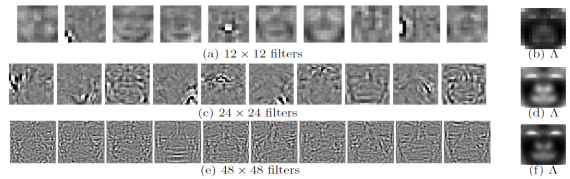
TIMIT Phone Recogntion



- MrDBN reduces error by 1.5%
- Obtaining core test error of 20.3%

Generative Experiments

Toronto Face Database

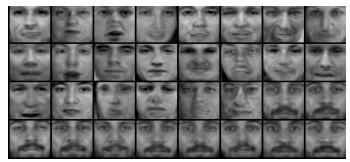


Learned filters at different frequencies

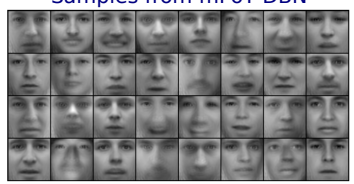
Learning the noise model is important for better generations

Figure 2: Filters and residual variances of all three resolutions of MrDBN trained on faces.

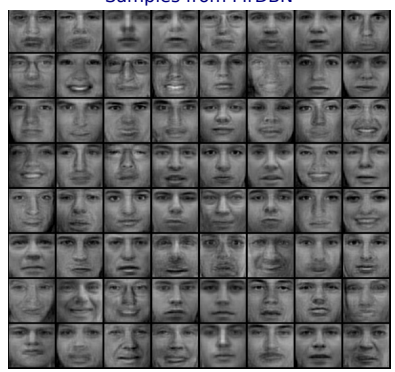
Samples from standard DBN



Samples from mPoT DBN



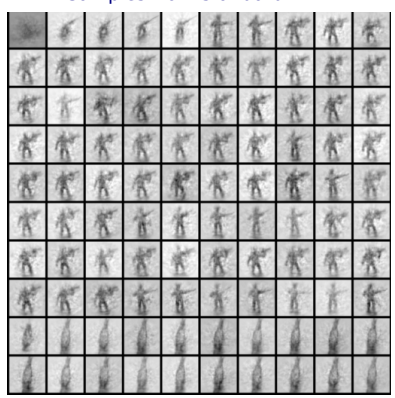
Samples from MrDBN



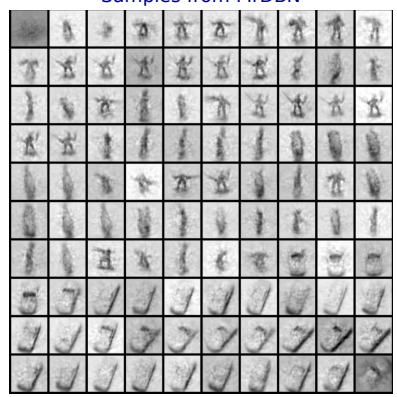
- Superior model of facial details
- 100 Gibbs iterations between samples

NORB

Samples from standard DBN



Samples from MrDBN



- Each sample generated after 1000 alternating Gibbs iterations
- Gibbs chain for MrDBN mixes faster