Unsupervised Learning of Hierarchical Models

Marc'Aurelio Ranzato - *Geoff Hinton*

in collaboration with Josh Susskind and Vlad Mnih

Advanced Machine Learning, 9 March 2011

Example: facial expression recognition

Unlabeled face images

Few labeled face images

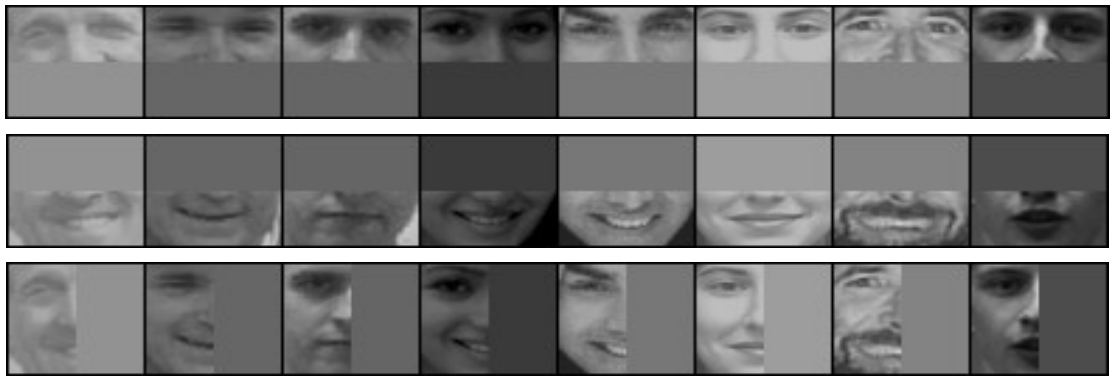
happiness

neutral

anger

fear

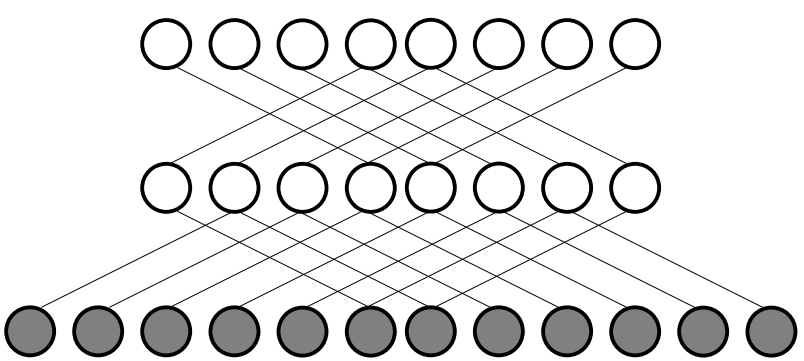
sadness



What is the expression?
(unseen occluded images)

Example: facial expression recognition

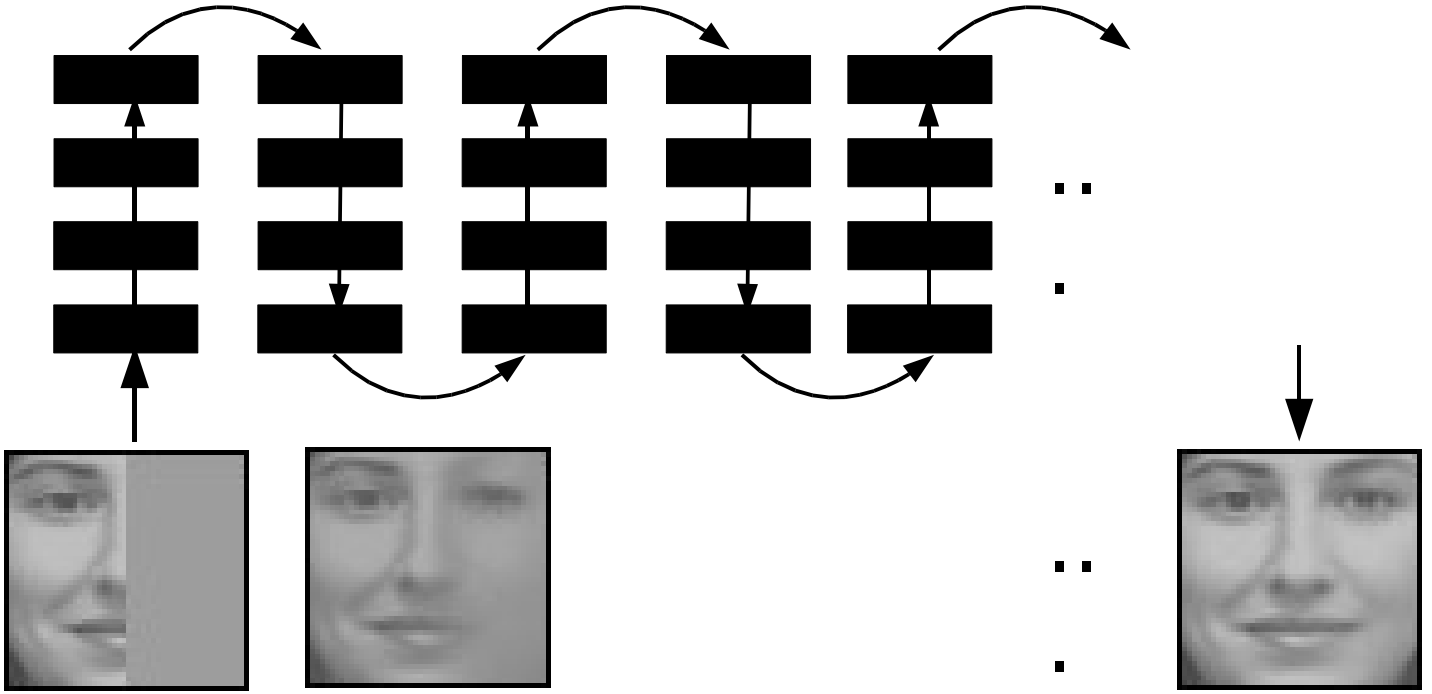
1) training on unlabeled data: maximum likelihood



Ranzato, Susskind, Mnih, Hinton "On deep generative models.." CVPR11

Example: facial expression recognition

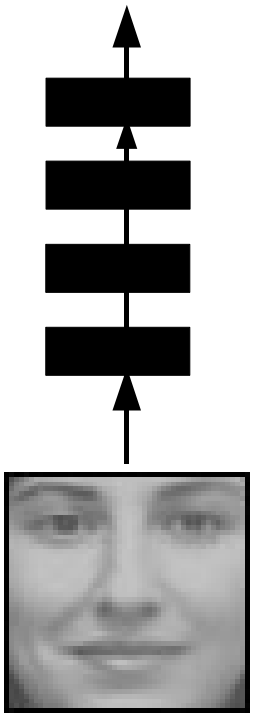
1) training on unlabeled data: maximum likelihood
2) use the generative model to inpute missing values



Ranzato, Susskind, Mnih, Hinton "On deep generative models.." CVPR11

Example: facial expression recognition

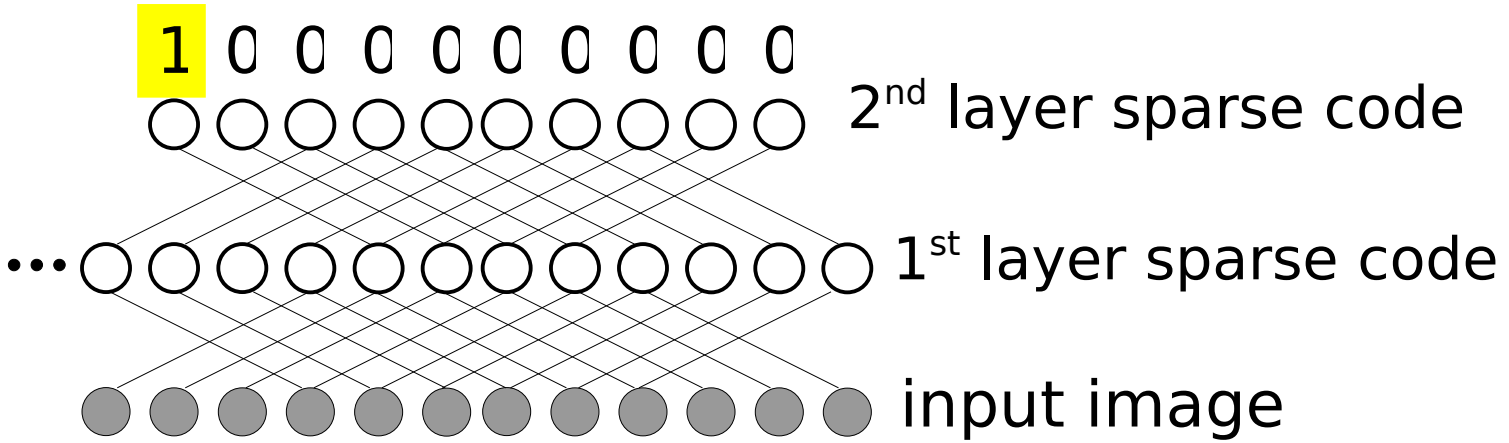
1) training on unlabeled data: maximum likelihood
2) use the generative model to inpute missing values
3) use the latent representation as features
4) train classifier on labeled features



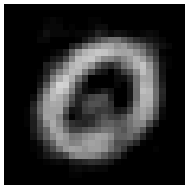
Ranzato, Susskind, Mnih, Hinton "On deep generative models.." CVPR11

The power of hierarchical representations

MNIST digits



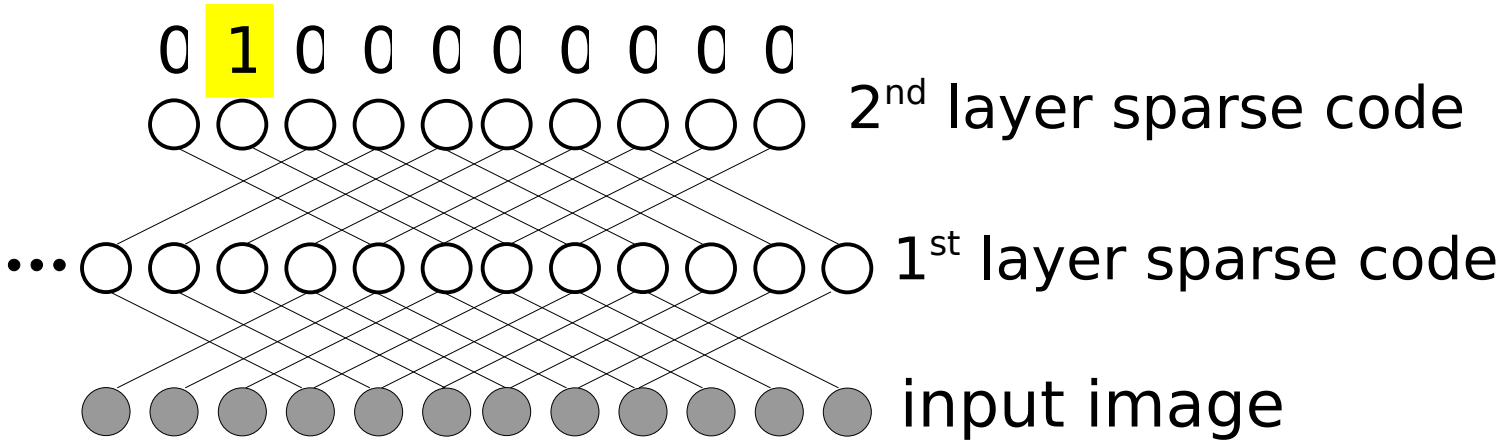
Receptive field of 2nd layer units



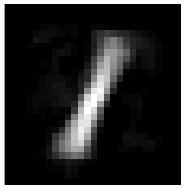
Ranzato et al. "Sparse feature learning for Deep Belief Nets" NIPS07

The power of hierarchical representations

MNIST digits



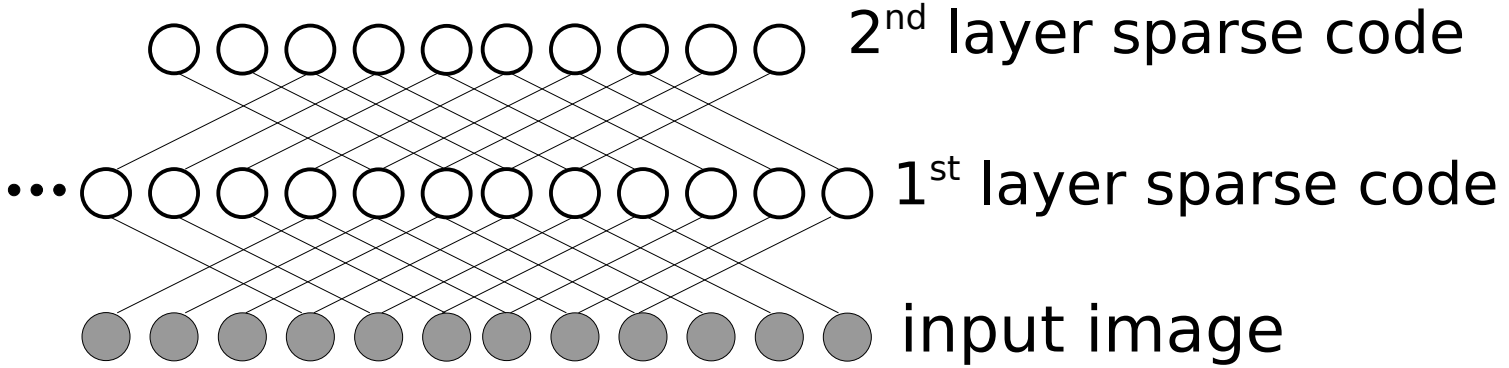
Receptive field of 2nd layer units



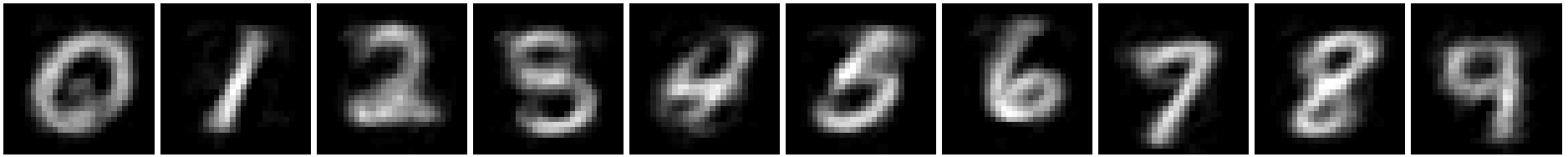
Ranzato et al. "Sparse feature learning for Deep Belief Nets" NIPS07

The power of hierarchical representations

MNIST digits



Receptive fields of 2nd layer units



Hierarchical representations can discover complex dependencies!

Ranzato et al. "Sparse feature learning for Deep Belief Nets" NIPS07

Unsupervised Learning

- we will learn hierarchical representations in a greedy fashion using unlabeled data (little if any assumption on the input)

- we can focus on single layer unsupervised algorihtms
    - how can we interpret unsupervised algorithms?
    - how can they learn representations?
    - how can we compare them?

Two Approaches to Unsupervised Learning

1st strategy: constrain latent representation &
optimize score only at training samples
- K-Means
- sparse coding

Ranzato et al. NIPS 06, Ranzato et al. CVPR 07, Ranzato et al. NIPS 07, Kavukcuoglu et al. CVPR 09

- use lower dimensional representations

Ranzato et al. ICML 08

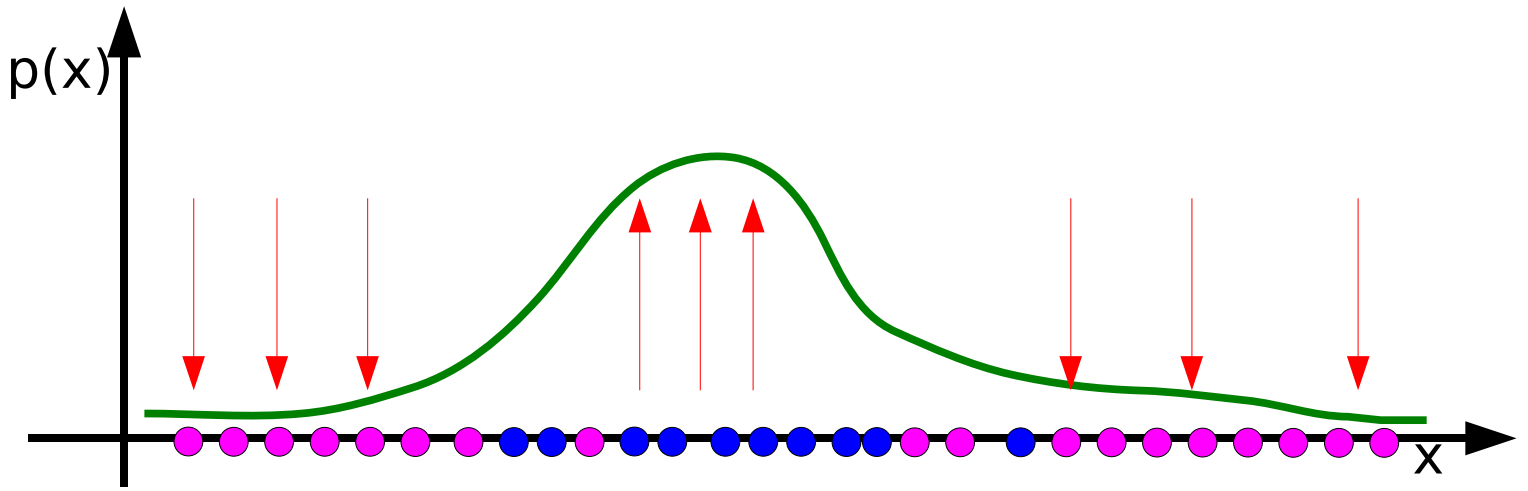
Two Approaches to Unsupervised Learning

1st strategy: constrain latent representation &
optimize score only at training samples
- K-Means
- sparse coding

- use lower dimensional representations

2nd strategy: optimize score for training samples while normalizing the score over the whole space (maximum likelihood)

Ranzato et al. AISTATS 10, Ranzato et al. CVPR 10, Ranzato et al. NIPS 10, Ranzato et al. CVPR 11



Two Approaches to Unsupervised Learning

	Pros	Cons
1st strategy (constrain code)	- efficient learning - defined through objective function (monitor training)	- choice of functions - not robust to missing inputs - not good for generation
2nd strategy (maximum likelihood)	- intuitive design - it can generate - compositionality	- hard to train (variational inference, MCMC, etc.)

Outline

- mathematical formulation of the model
- training
- learning acoustic features for spech recognition
- generation of natural images
- recognition of facial expression under occlusion
- conclusion

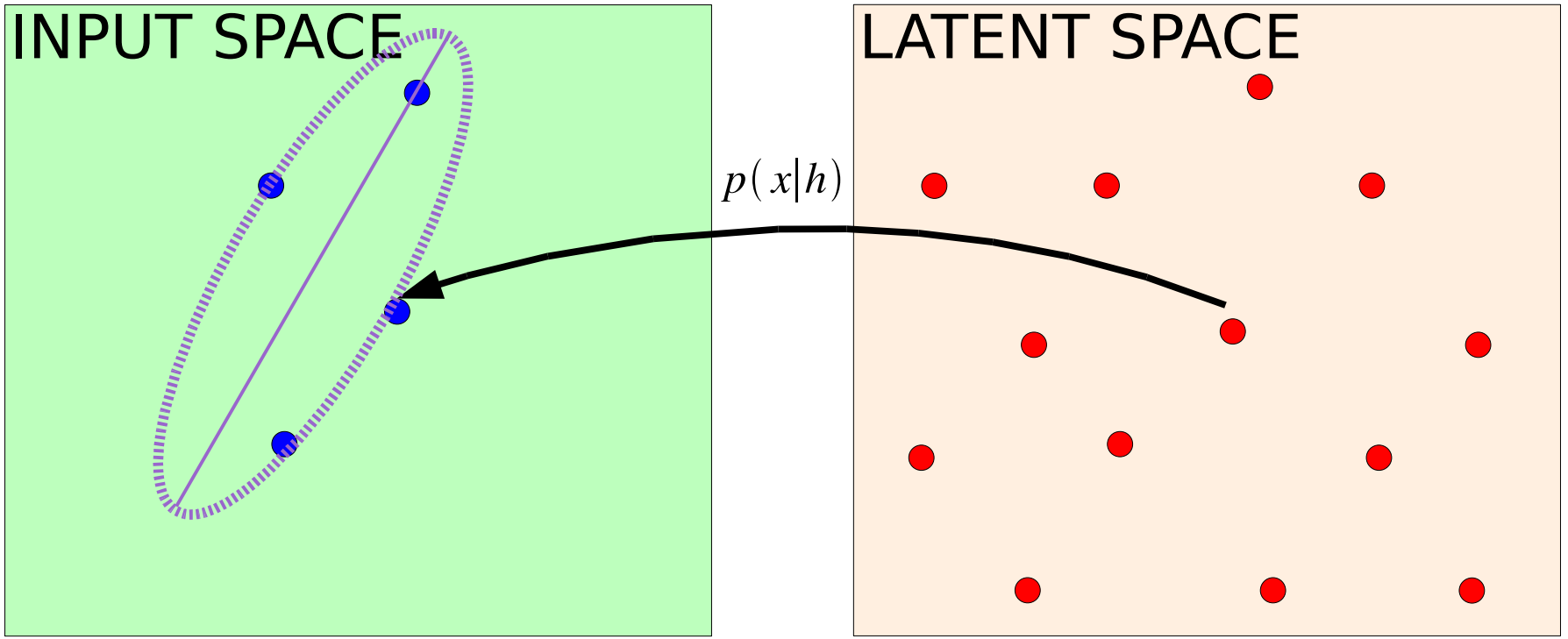
Outline

- mathematical formulation of the model
- training
- learning acoustic features for spech recognition
- generation of natural images
- recognition of facial expression under occlusion
- conclusion

Means and Covariances: PPCA

$$p(x|h) \neq p(x)$$

$p(x)$ specifies covariance <u>across all data points</u>

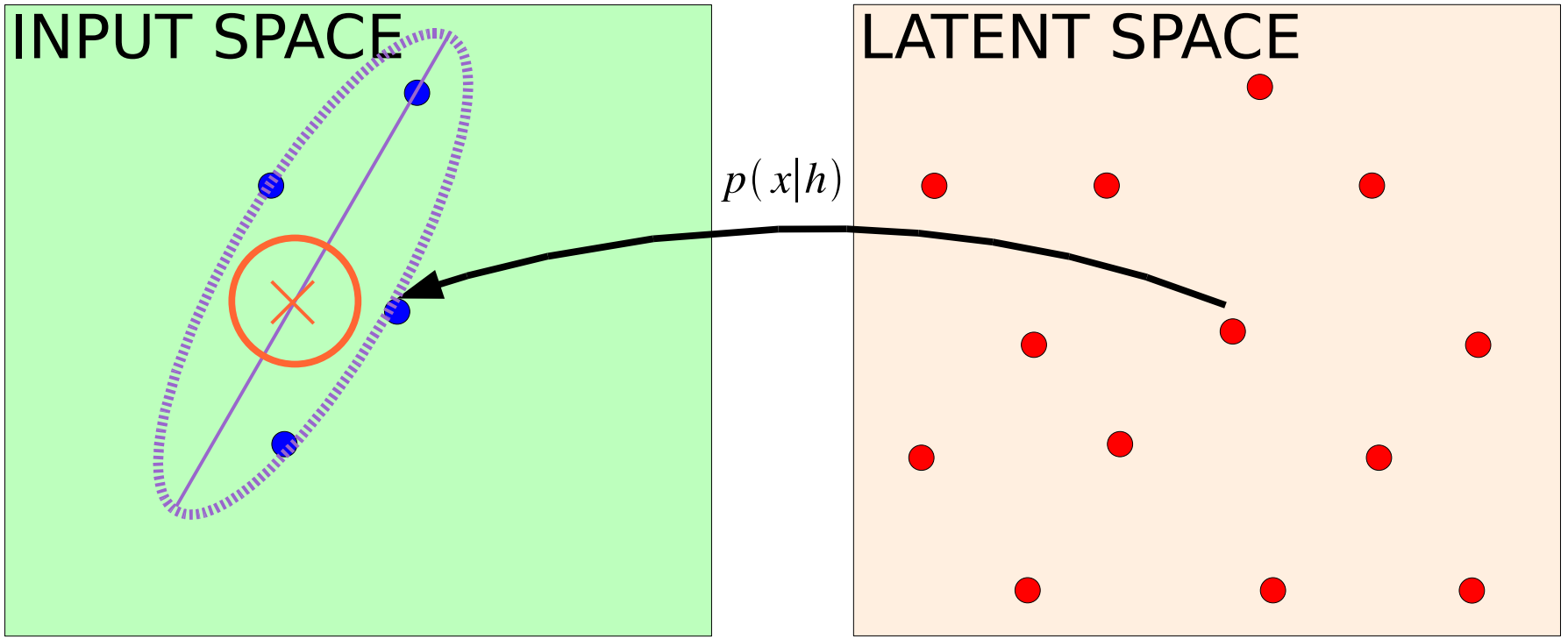


Means and Covariances: PPCA

$$p(x|h) \neq p(x)$$

$h_t \sim p(h|x_t)$

$p(x|h_t)$ specifies distribution <u>for each "data point"</u>



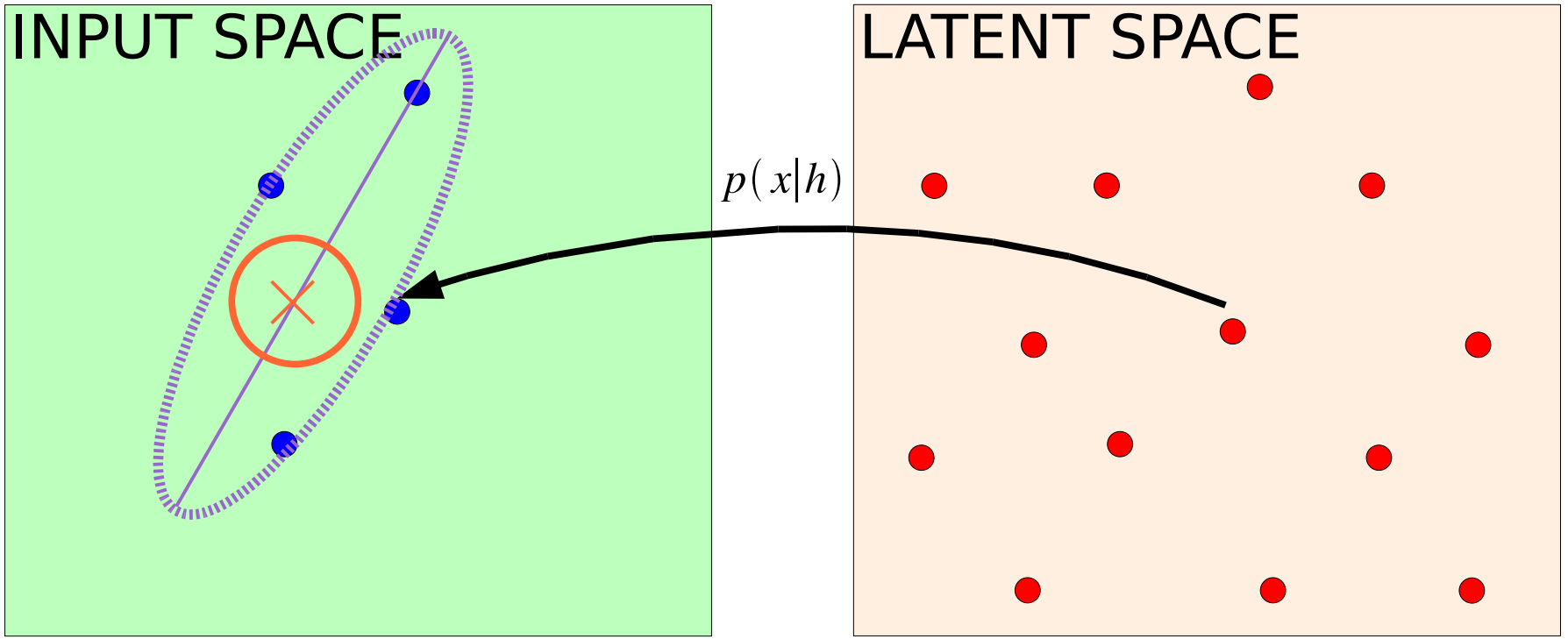
● **Training sample**

● **Latent vector**

Means and Covariances: PPCA

$$p(x) = \int_h p(x|h)\, p(h)$$

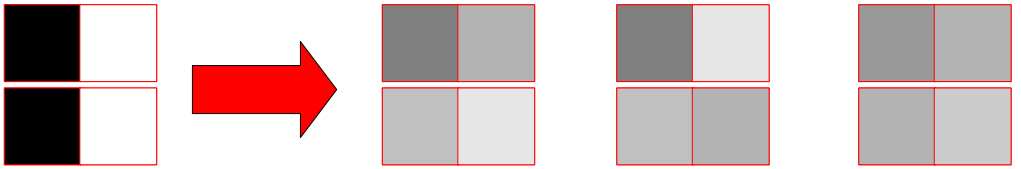
$p(x|h)$ specifies distribution <u>for each "data point"</u>

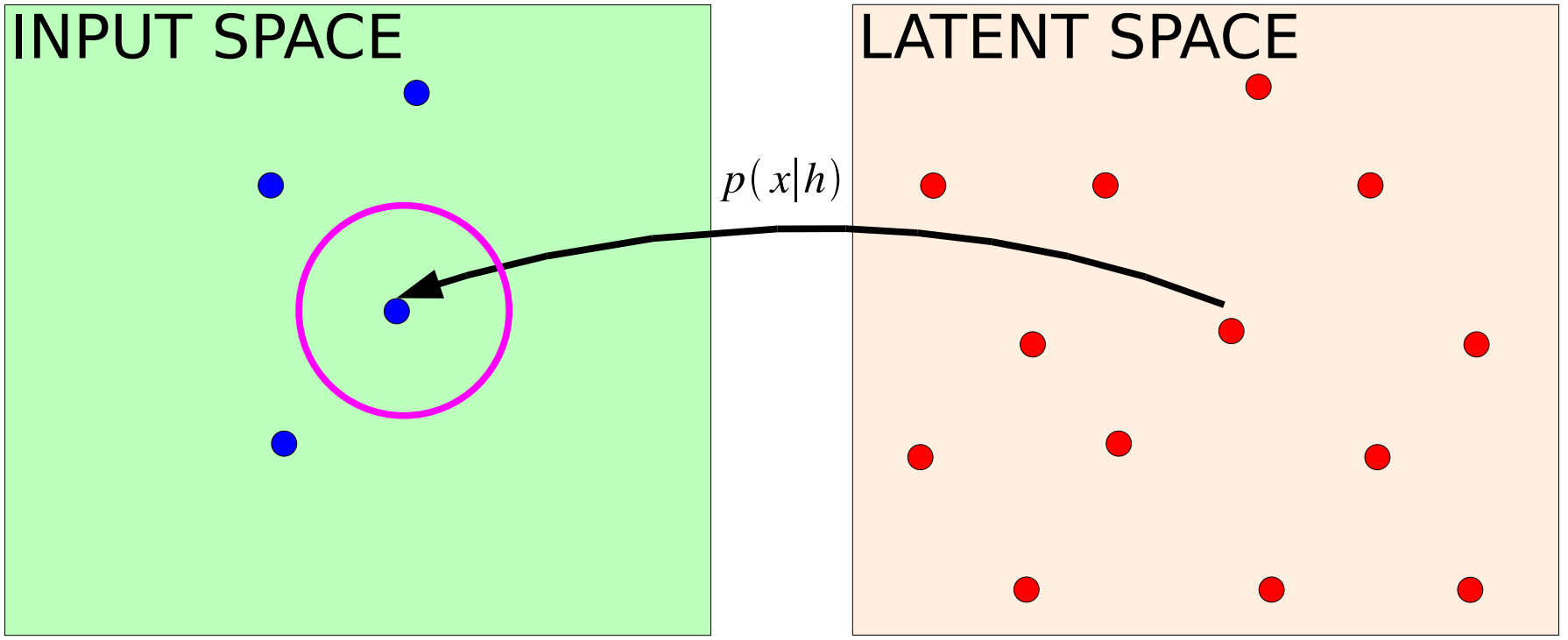


Conditional Distribution Over Input

$$p(x|h) = N(mean(h), D)$$

- examples: PPCA, Factor Analysis, ICA, Gaussian RBM





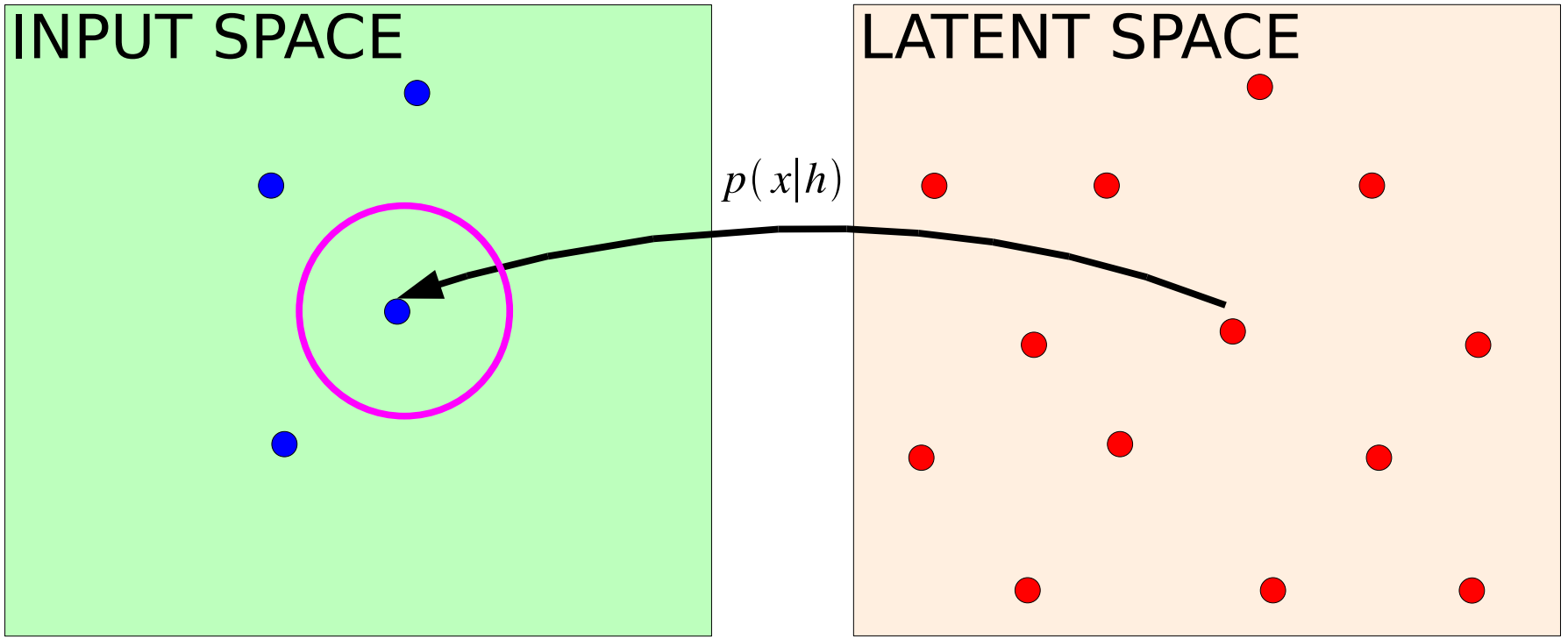
Training sample **Latent vector**

Conditional Distribution Over Input

$$p(x|h) = N(mean(h), D)$$

- examples: PPCA, Factor Analysis, ICA, Gaussian RBM

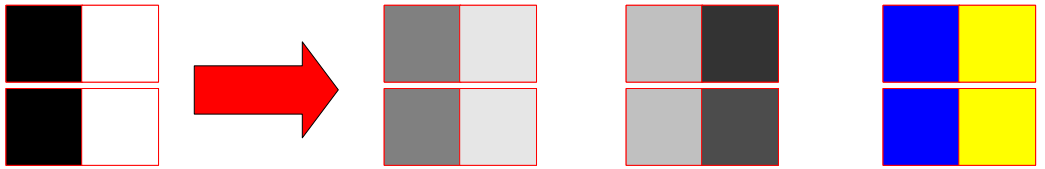
Poor model of dependency between input variables

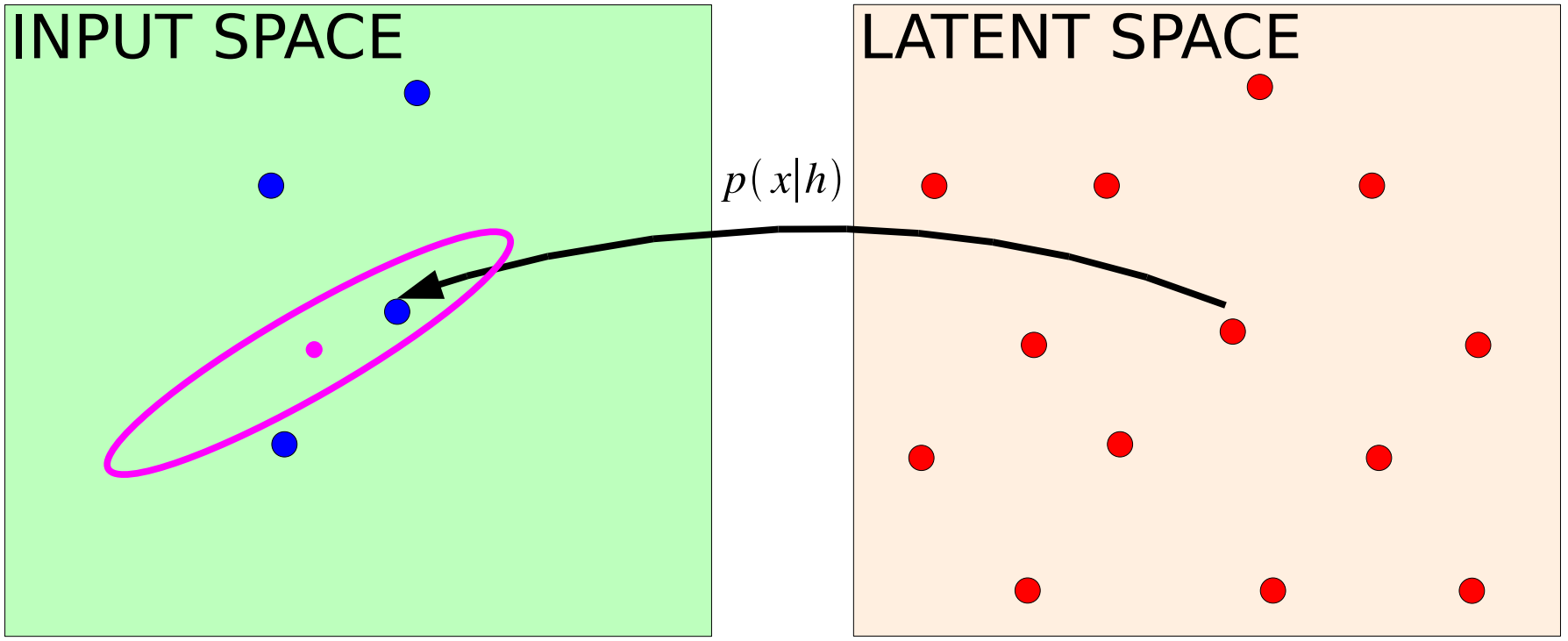


Conditional Distribution Over Input

$$p(x|h) = N(0, Covariance(h))$$

- examples: PoT, cRBM





Training sample **Latent vector**

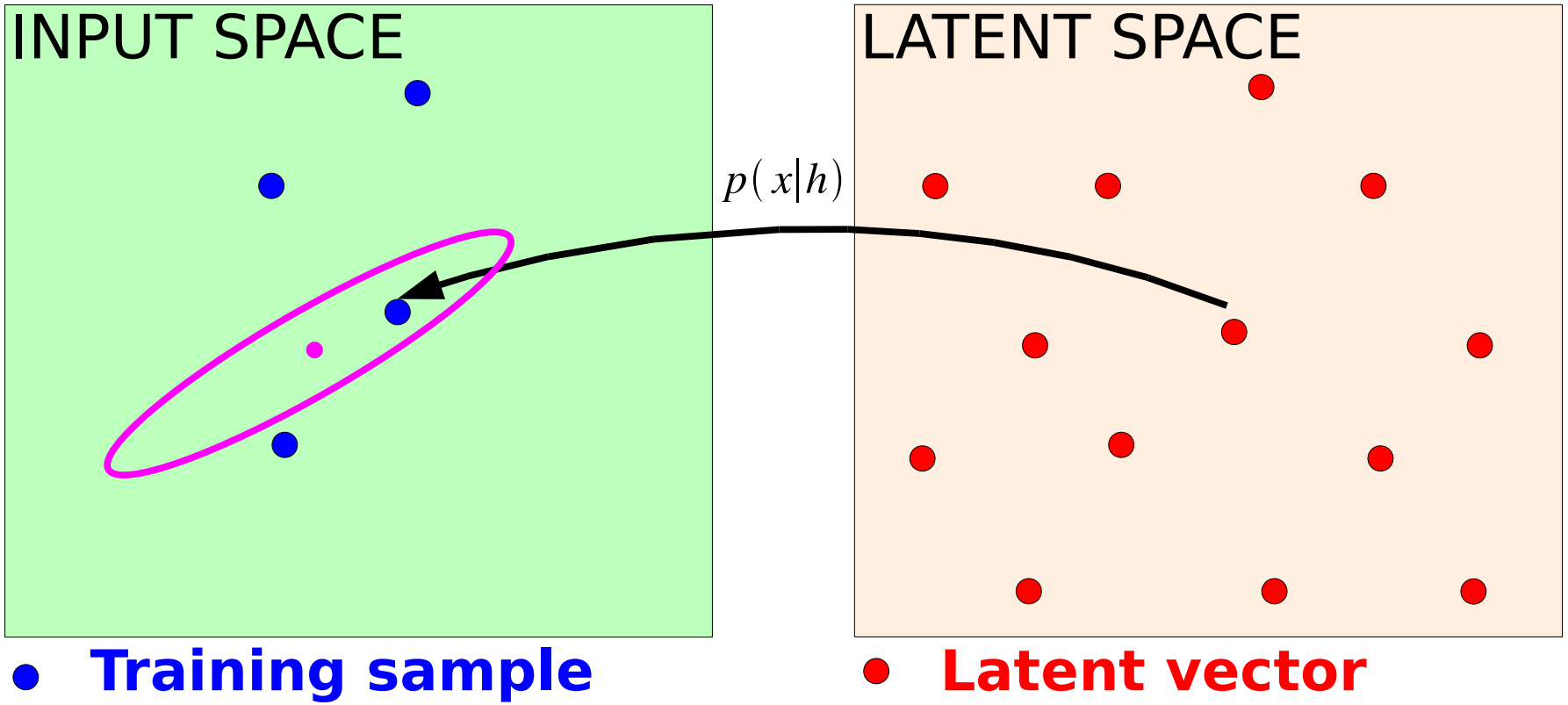
Welling et al. NIPS 2003, Ranzato et al. AISTATS 10

Conditional Distribution Over Input

$$p(x|h) = N(0, Covariance(h))$$

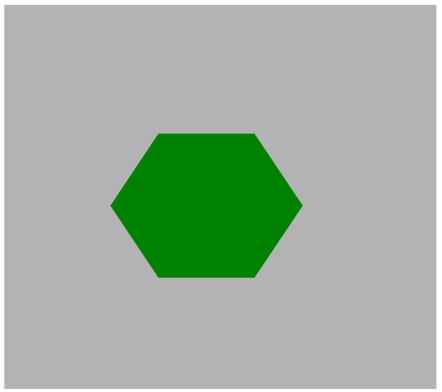
- examples: PoT, cRBM

Poor model of mean intensity

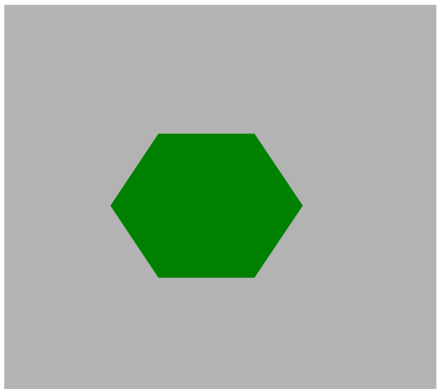


Welling et al. NIPS 2003, Ranzato et al. AISTATS 10

input image

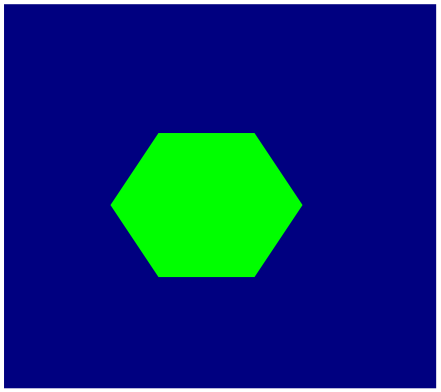


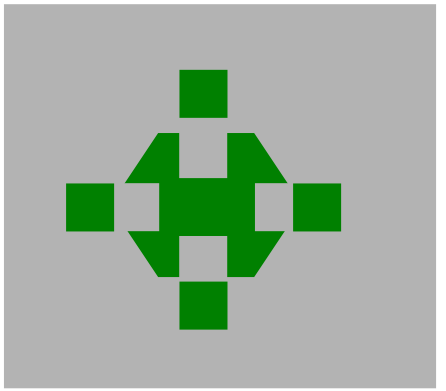
input image



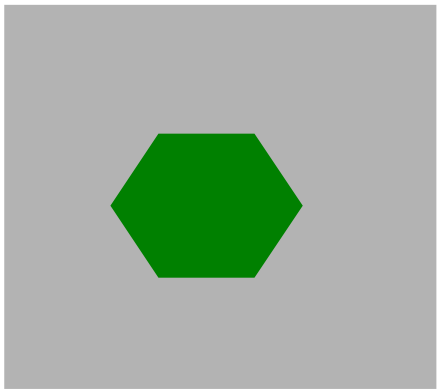
Which image has similar mean (but different covariance)?

Which image has same covariance (but different mean)?

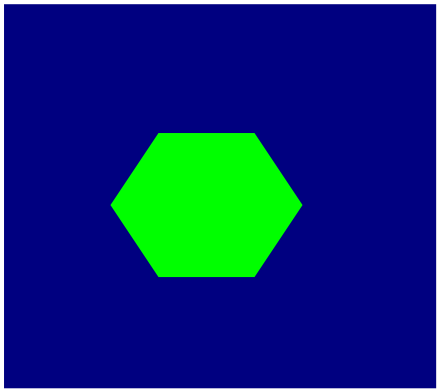




input image

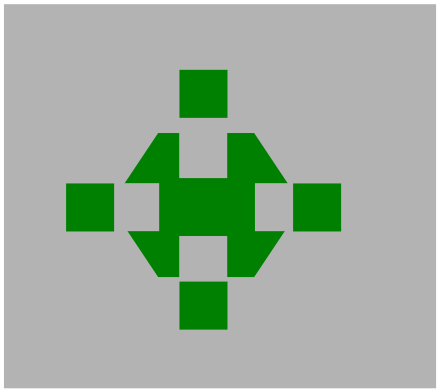


same covariance
but different mean



Equivalent image according to PoT

similar mean
but different covariance

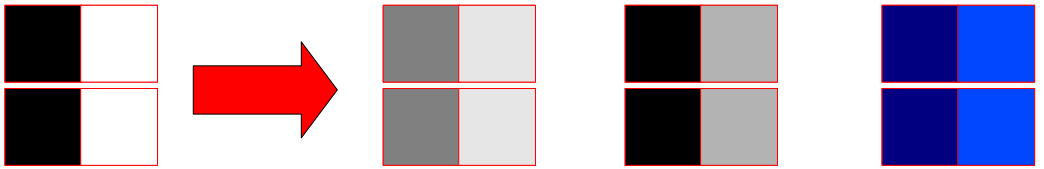


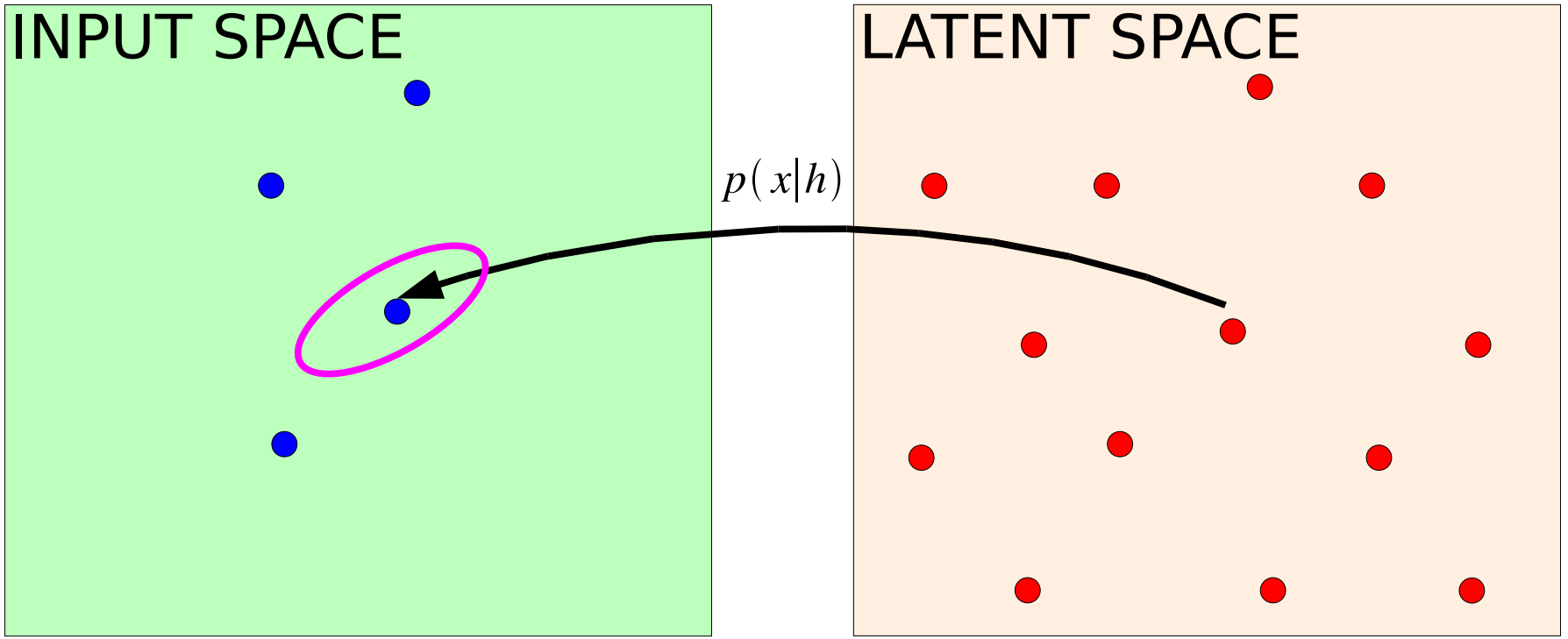
Equivalent image according to FA

Conditional Distribution Over Input

$$p(x|h) = N(mean(h), Covariance(h))$$

- this is what we propose: mcRBM, mPoT



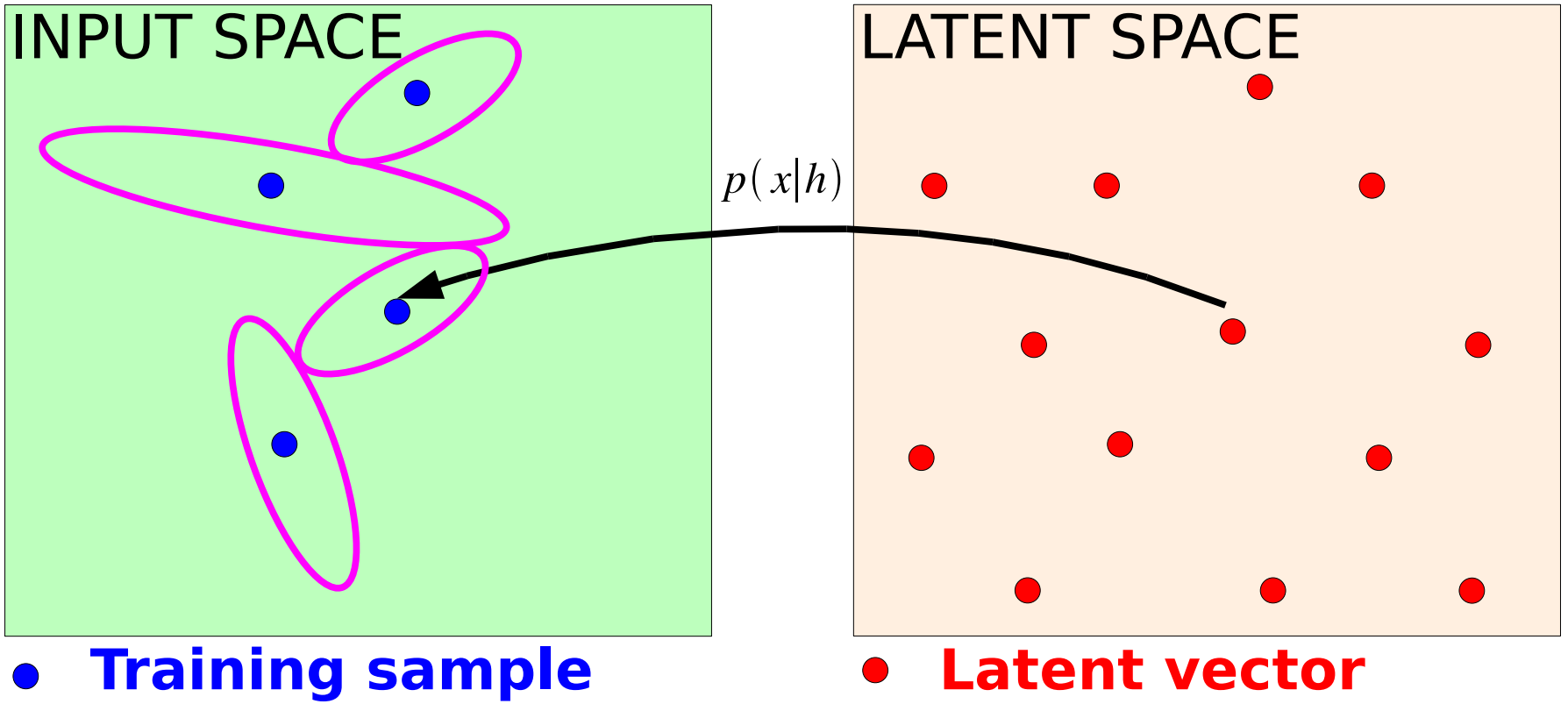


Ranzato et al. CVPR 10, Ranzato et al. NIPS 2010, Ranzato et al. CVPR 11

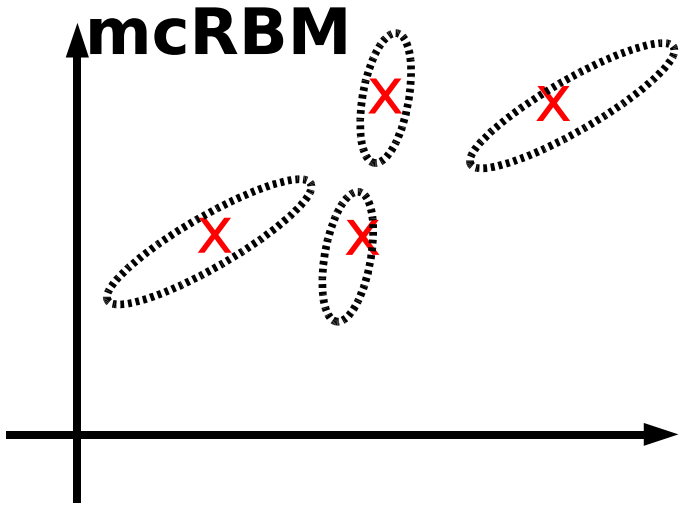
Conditional Distribution Over Input

$$p(x|h) = N(mean(h), Covariance(h))$$

- this is what we propose: mcRBM, mPoT



Ranzato et al. CVPR 10, Ranzato et al. NIPS 2010, Ranzato et al. CVPR 11



- two sets of latent variables to modulate mean and covariance of the conditional distribution over the input

- energy-based model

$$p(x, h^m, h^c) \propto \exp(-E(x, h^m, h^c))$$

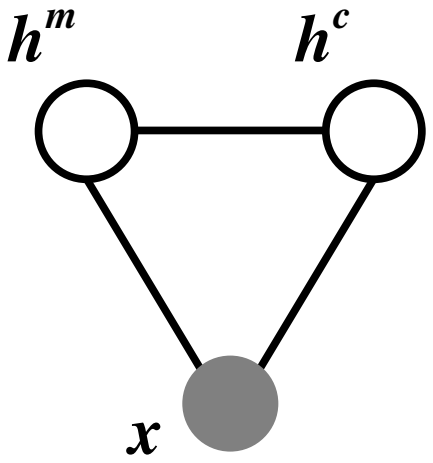
$$x \in \mathbb{R}^D$$
$$h^c \in \{0,1\}^M$$
$$h^m \in \{0,1\}^N$$

Ranzato Hinton CVPR 10

- easy generation
- slow inference

$$p(x|h^m, h^c) = N(m(h^m), \Sigma(h^c))$$

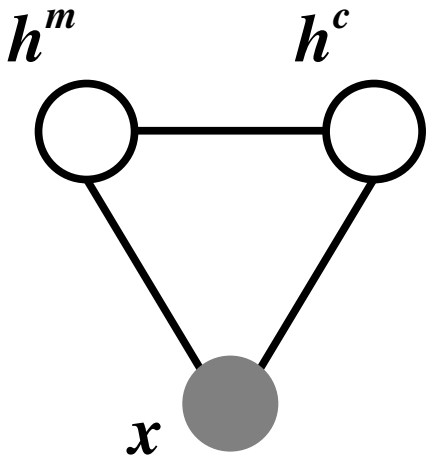
$$E = \frac{1}{2}(x-m)' \Sigma^{-1}(x-m)$$



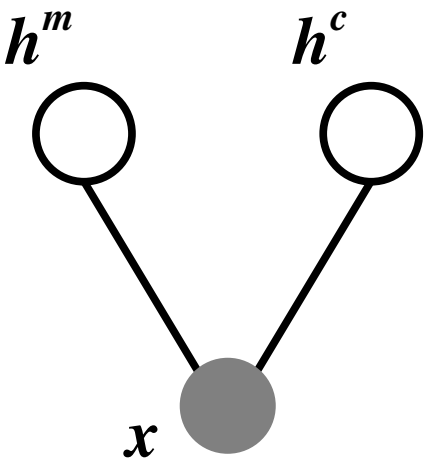
- easy generation
- slow inference

$$p(x|h^m, h^c) = N(m(h^m), \Sigma(h^c))$$

$$E = \frac{1}{2}(x-m)' \Sigma^{-1}(x-m)$$



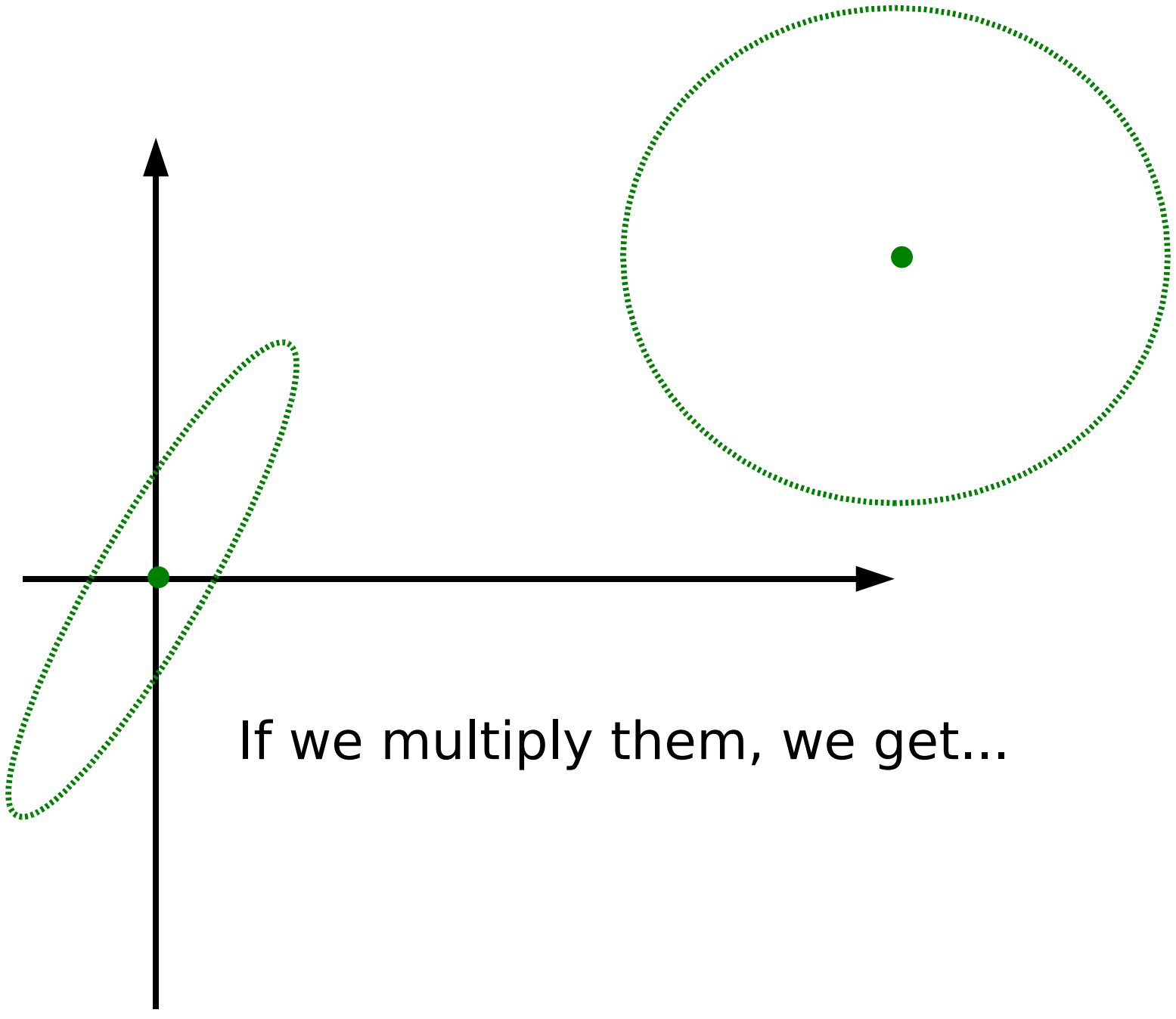
- less easy generation
- fast inference



$$E = \frac{1}{2} x' \Sigma^{-1} x - m x$$

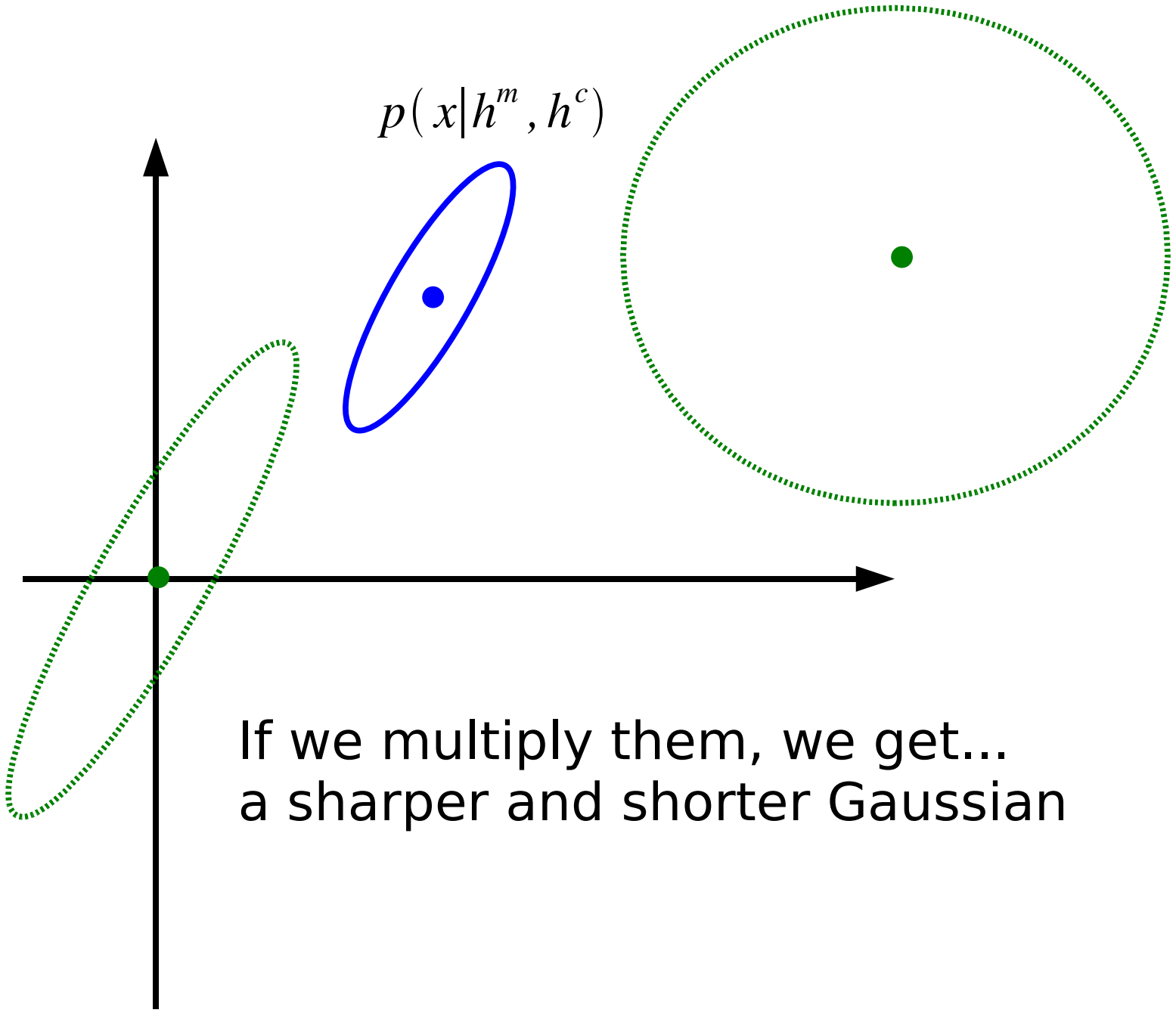
$$p(x|h^m, h^c) = N(\Sigma(h^c)^{-1} m(h^m), \Sigma(h^c))$$

Geometric interpretation



If we multiply them, we get...

Geometric interpretation



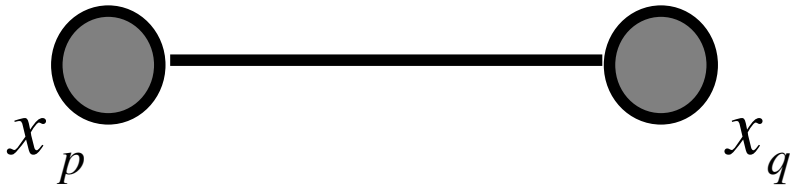
If we multiply them, we get...
a sharper and shorter Gaussian

Covariance part of the energy function:

$$E(x, h^c, h^m) = \frac{1}{2} x' \Sigma^{-1} x$$

$$x \in \mathbb{R}^D$$

$$\Sigma^{-1} \in \mathbb{R}^{D \times D}$$



pair-wise MRF

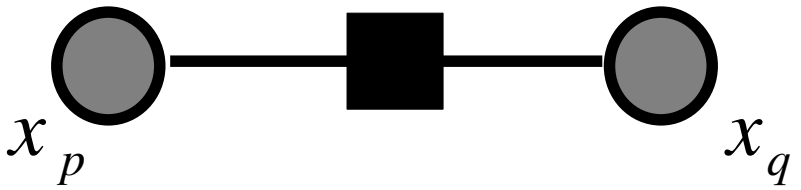
Covariance part of the energy function:

$$E(x, h^c, h^m) = \frac{1}{2} x' C C' x$$

factorization

$x \in \mathbb{R}^D$

$C \in \mathbb{R}^{D \times F}$



pair-wise MRF

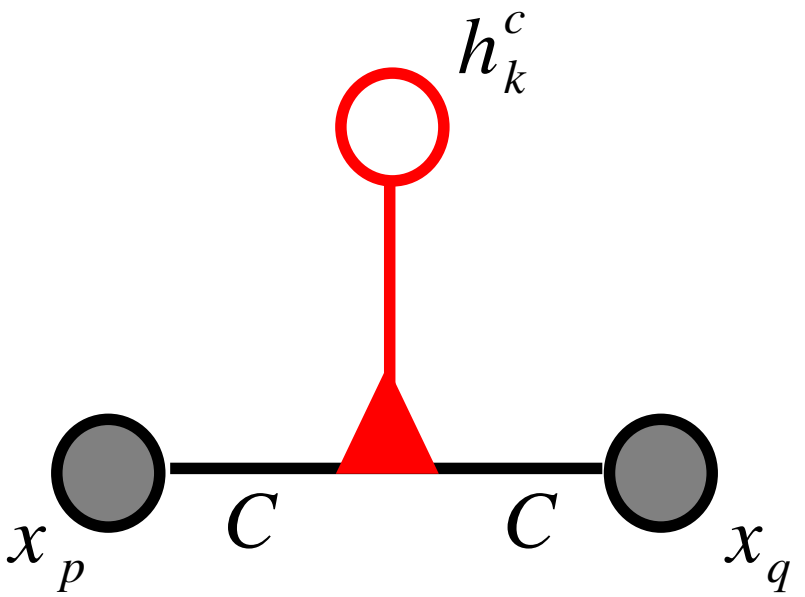
Covariance part of the energy function:

$$E(x, h^c, h^m) = \frac{1}{2} x' C[diag(h^c)] C' x$$

$x \in \mathbb{R}^D$ factorization + hiddens

$C \in \mathbb{R}^{D \times F}$

$h^c \in \{0,1\}^F$



gated MRF

Covariance part of the energy function:

$$E(x, h^c, h^m) = \frac{1}{2} x' C[diag(P h^c)] C' x$$

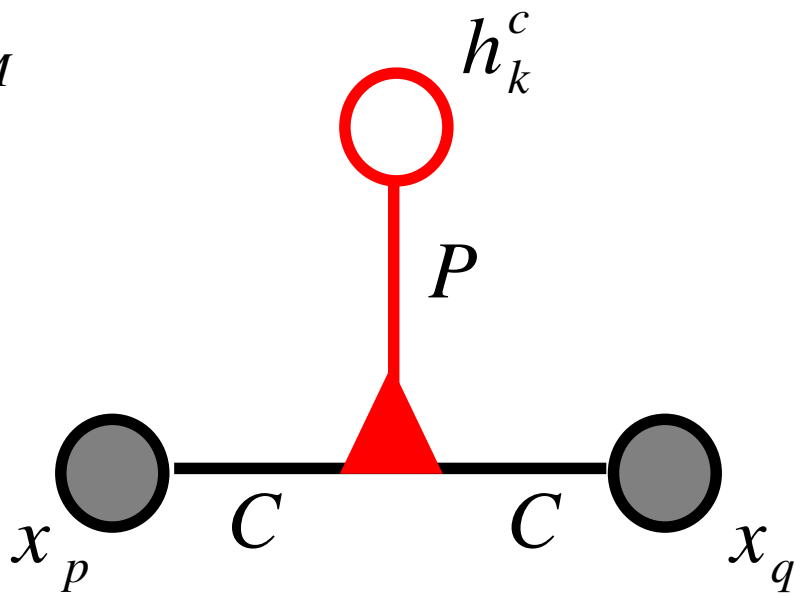
factorization + hiddens

$x \in \mathbb{R}^D$

$C \in \mathbb{R}^{D \times F}$

$h^c \in \{0,1\}^M$

$P \in \mathbb{R}^{F \times M}$



gated MRF

Overall energy function:

$$E(x, h^c, h^m) = \frac{1}{2} x' C [diag(P h^c)] C' x + \frac{1}{2} x' x - x' W h^m$$

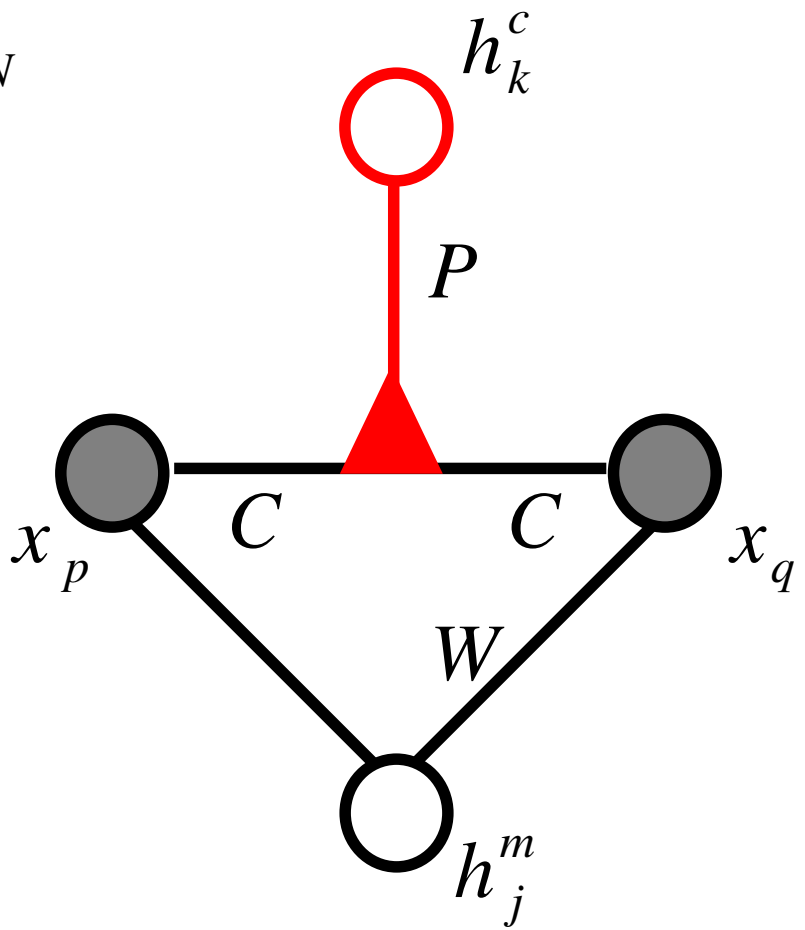
covariance part

mean part

$x \in \mathbb{R}^D$

$W \in \mathbb{R}^{D \times N}$

$h^m \in \{0,1\}^N$



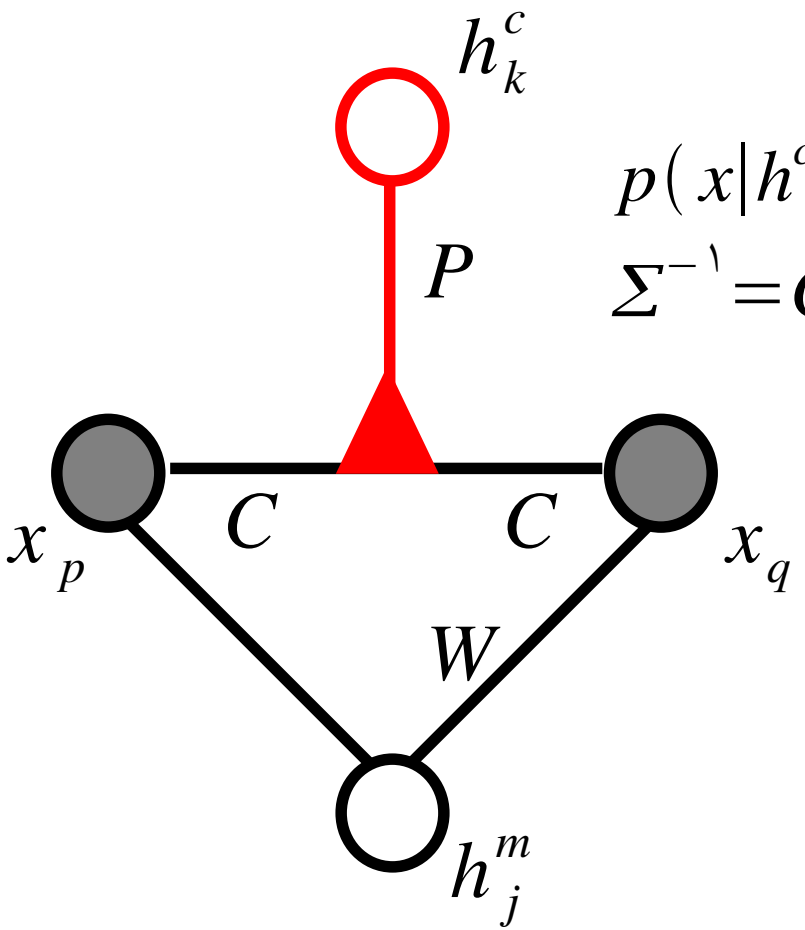
gated MRF

Overall energy function:

$$E(x, h^c, h^m) = \frac{1}{2} x' C [diag(P h^c)] C' x + \frac{1}{2} x' x - x' W h^m$$

covariance part

mean part



$$p(x|h^c, h^m) = N(\Sigma(Wh^m), \Sigma)$$

$$\Sigma^{-1} = C\, diag[P h^c] C' + I$$

gated MRF

Overall energy function:

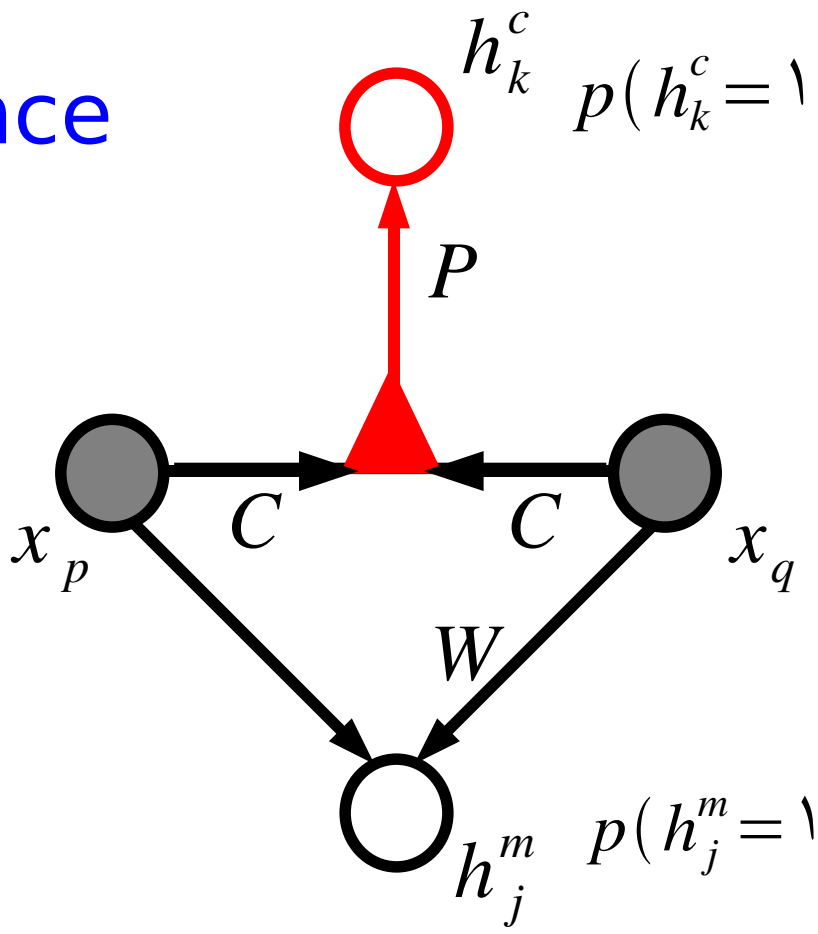
$$E(x, h^c, h^m) = \frac{1}{2} x' C [diag(P h^c)] C' x + \frac{1}{2} x' x - x' W h^m$$

covariance part

mean part

inference

$$p(h_k^c = 1 | x) = \sigma(-\frac{1}{2} P_k (C' x)^2 + b_k)$$



gated MRF

$$p(h_j^m = 1 | x) = \sigma(W_j' x + b_j)$$

Ranzato Hinton CVPR 10

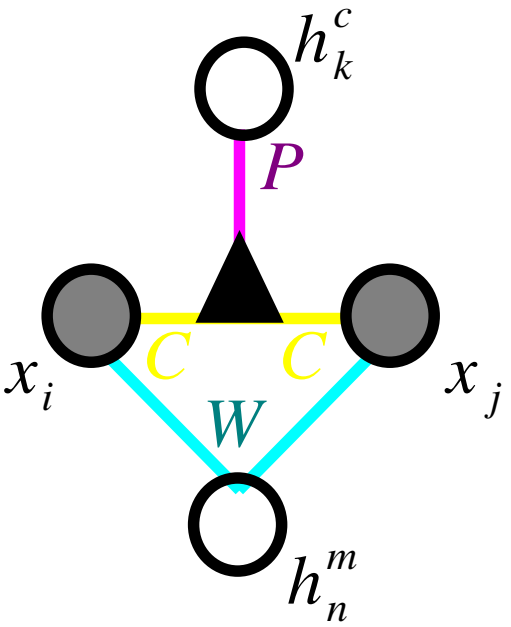
Outline

- mathematical formulation of the model
- training
- learning acoustic features for spech recognition
- generation of natural images
- recognition of facial expression under occlusion
- conclusion

Learning

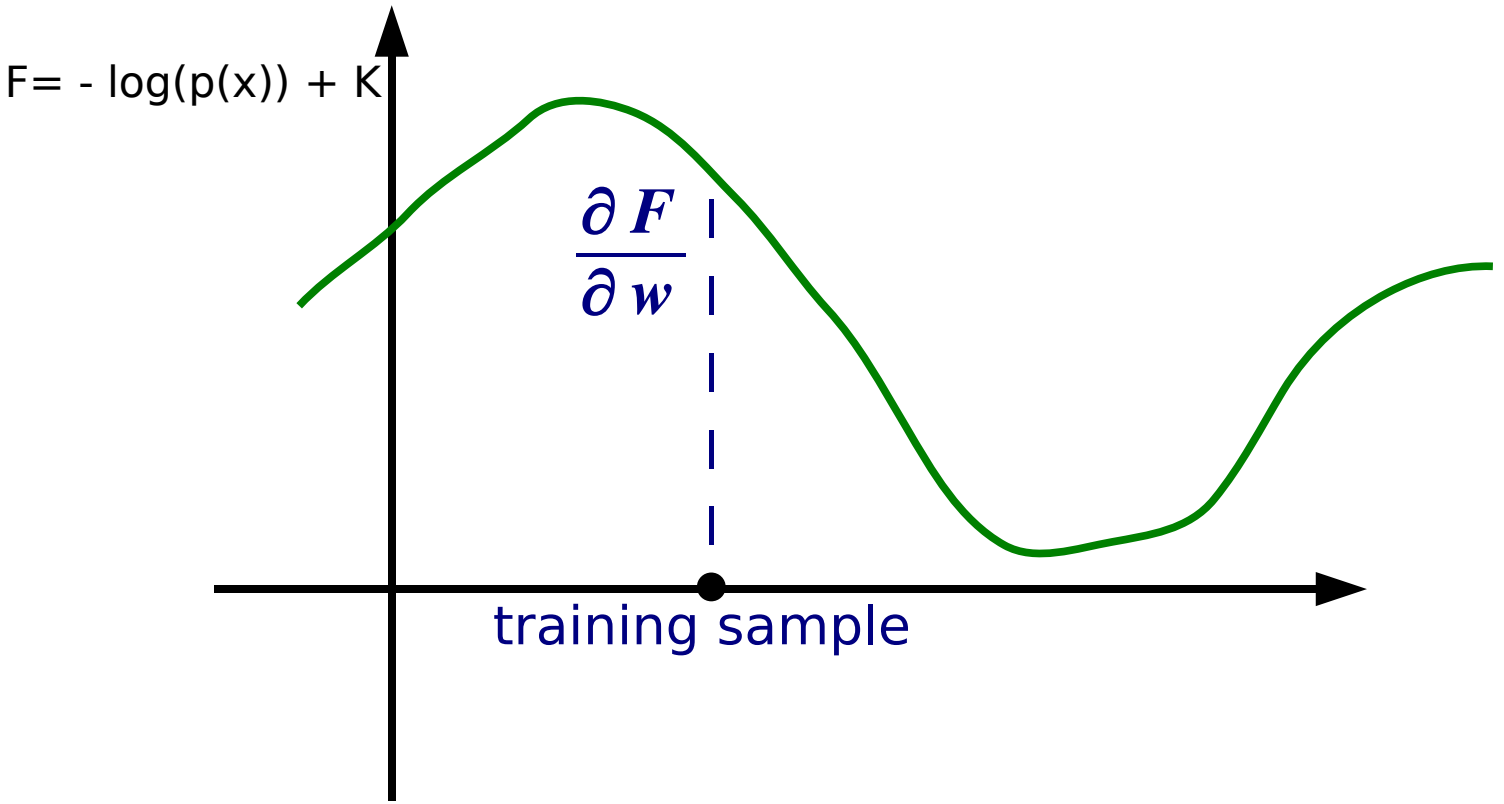
- maximum likelihood $p(x)=\frac{\int_{h^m,h^c} e^{-E(x,h^m,h^c)}}{\int_{x,h^m,h^c} e^{-E(x,h^m,h^c)}}$

  - Fast Persistent Contrastive Divergence
  - Hybrid Monte Carlo to draw samples

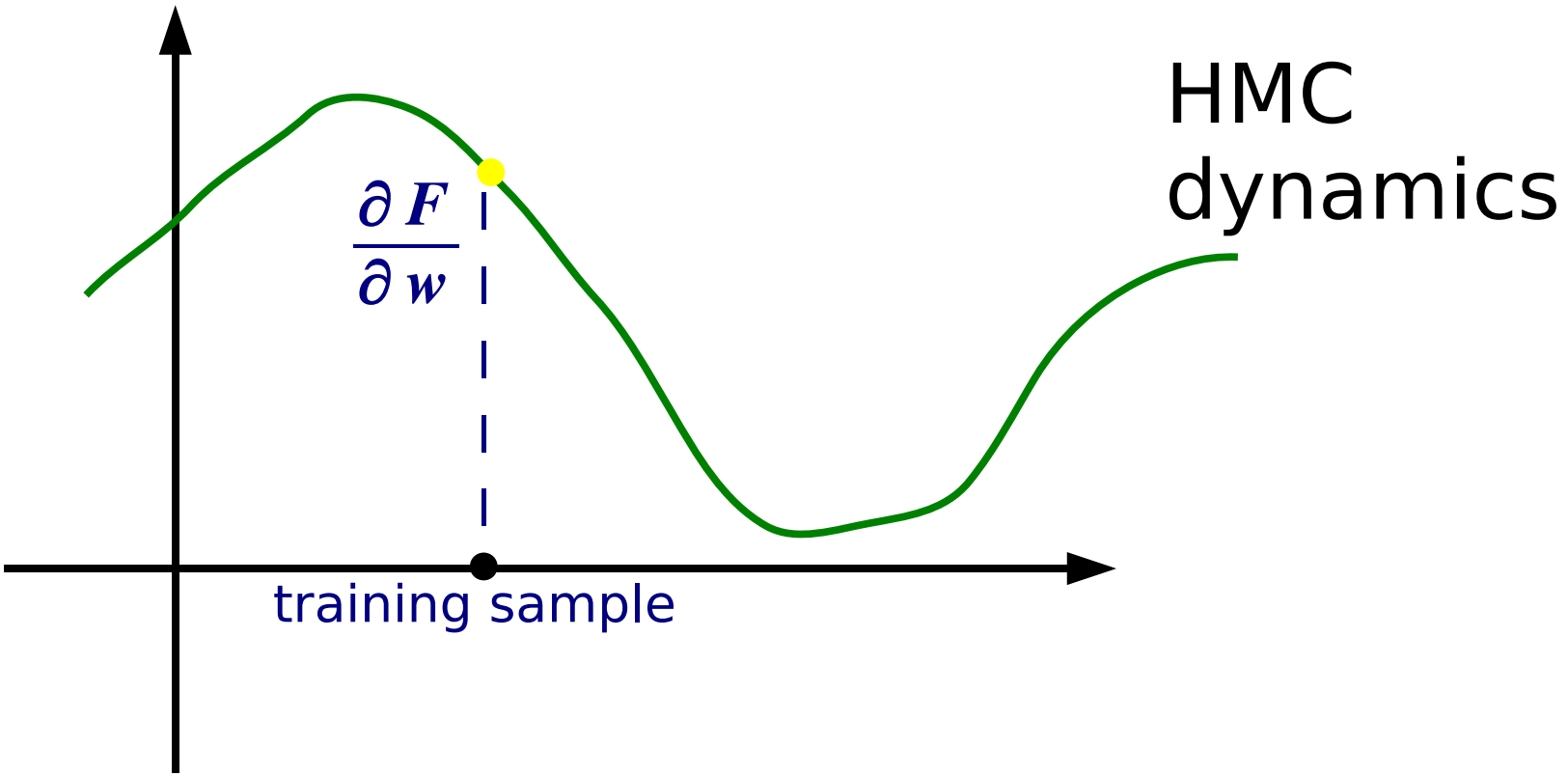


$$E=\frac{1}{2}x'C[diag(Ph^c)]C'x-x'Wh^m+\ldots$$

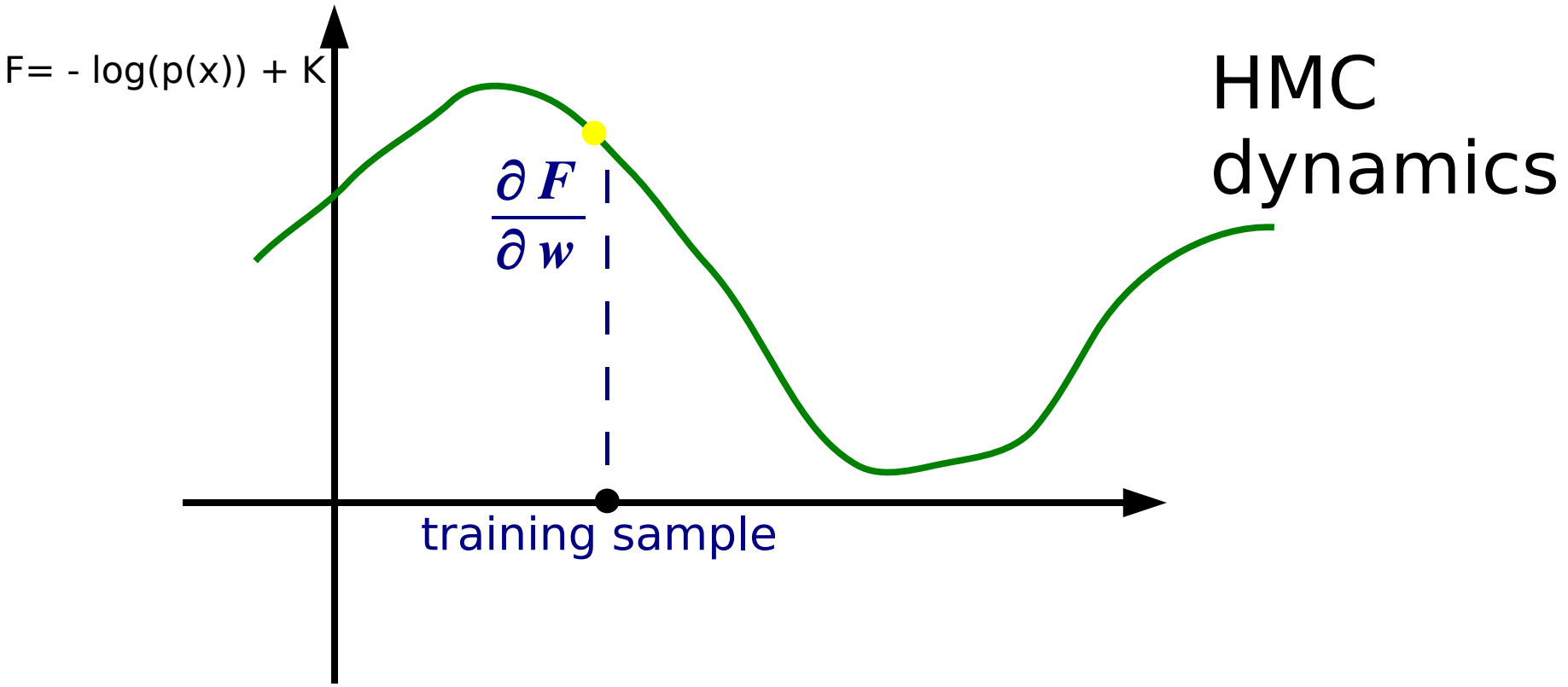
Learning



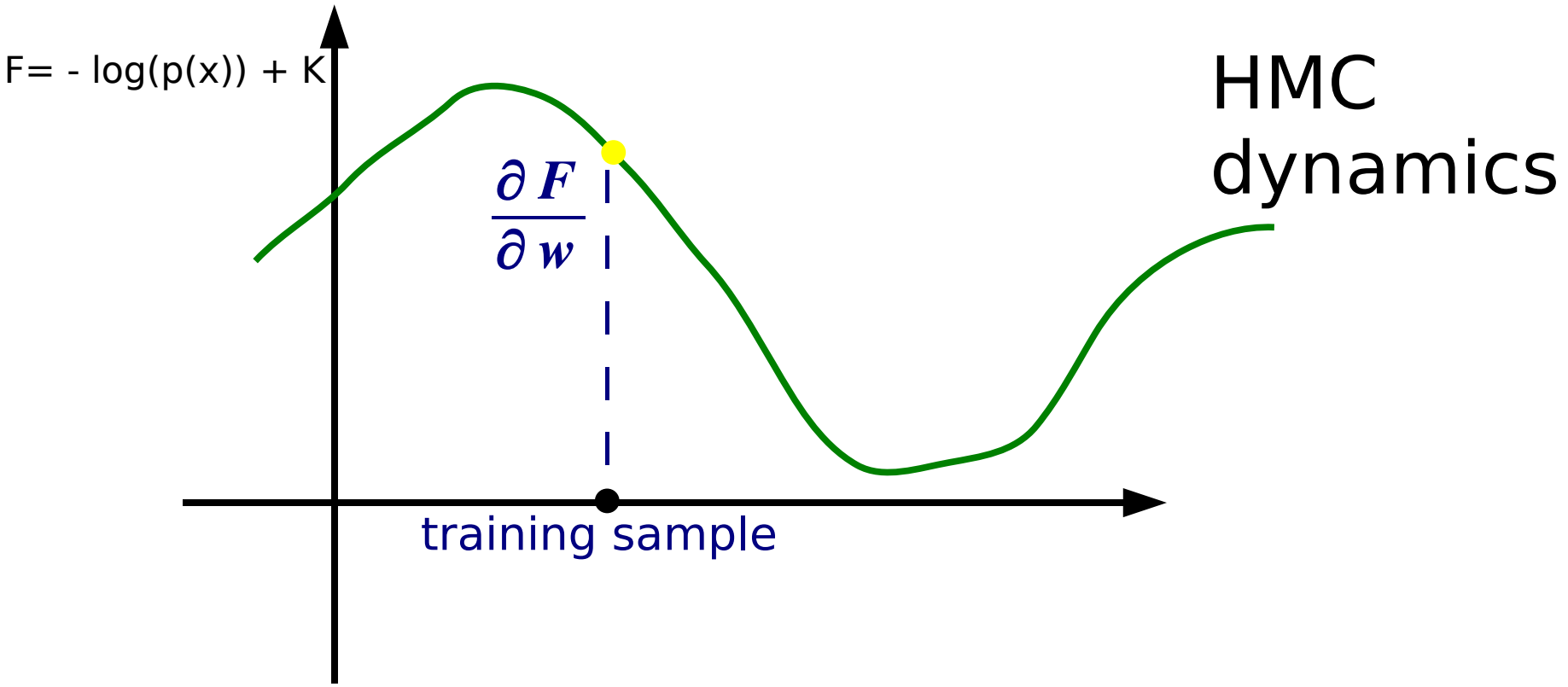
Learning



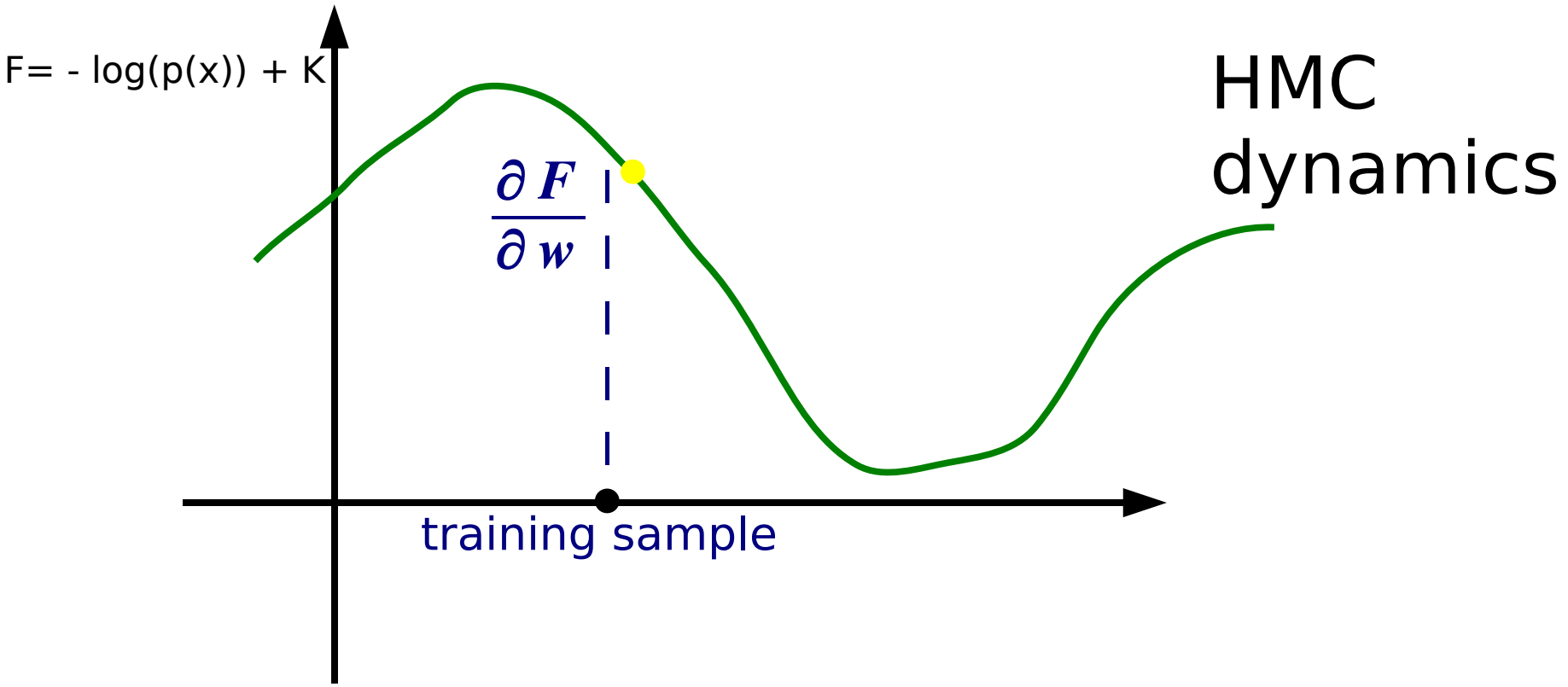
Learning


F= - log(p(x)) + K
$\frac{\partial F}{\partial w}$
HMC
dynamics
training sample

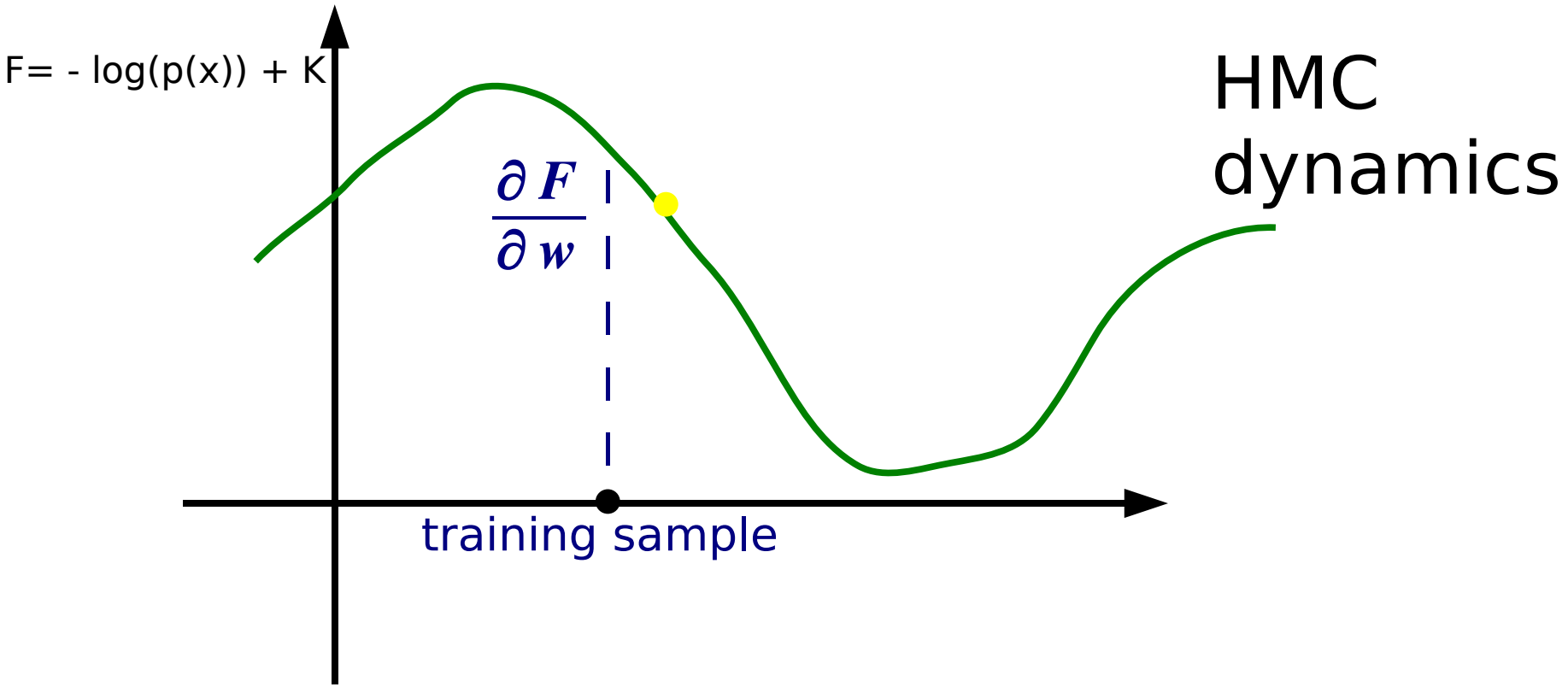
Learning


F= - log(p(x)) + K
$\frac{\partial F}{\partial w}$
HMC
dynamics
training sample

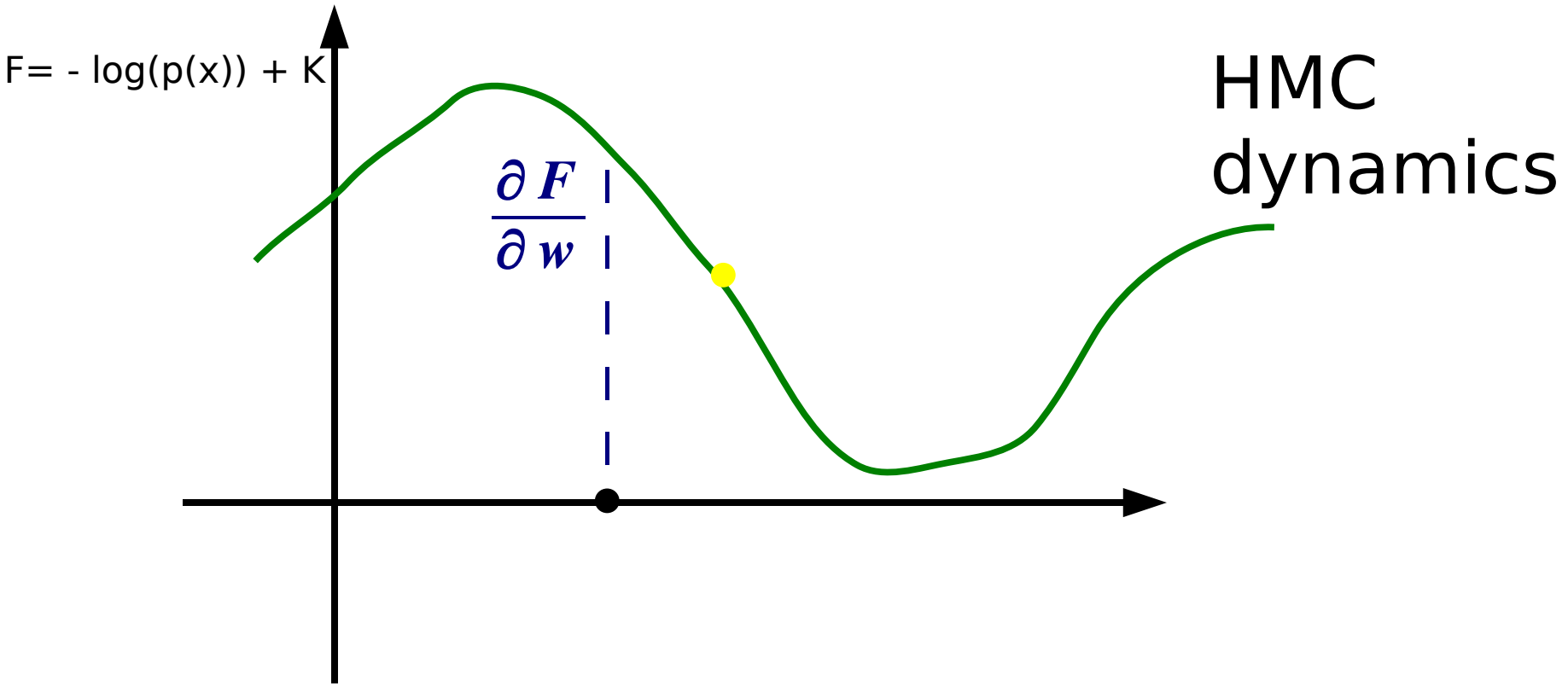
Learning


F= - log(p(x)) + K
$\frac{\partial F}{\partial w}$
HMC
dynamics
training sample

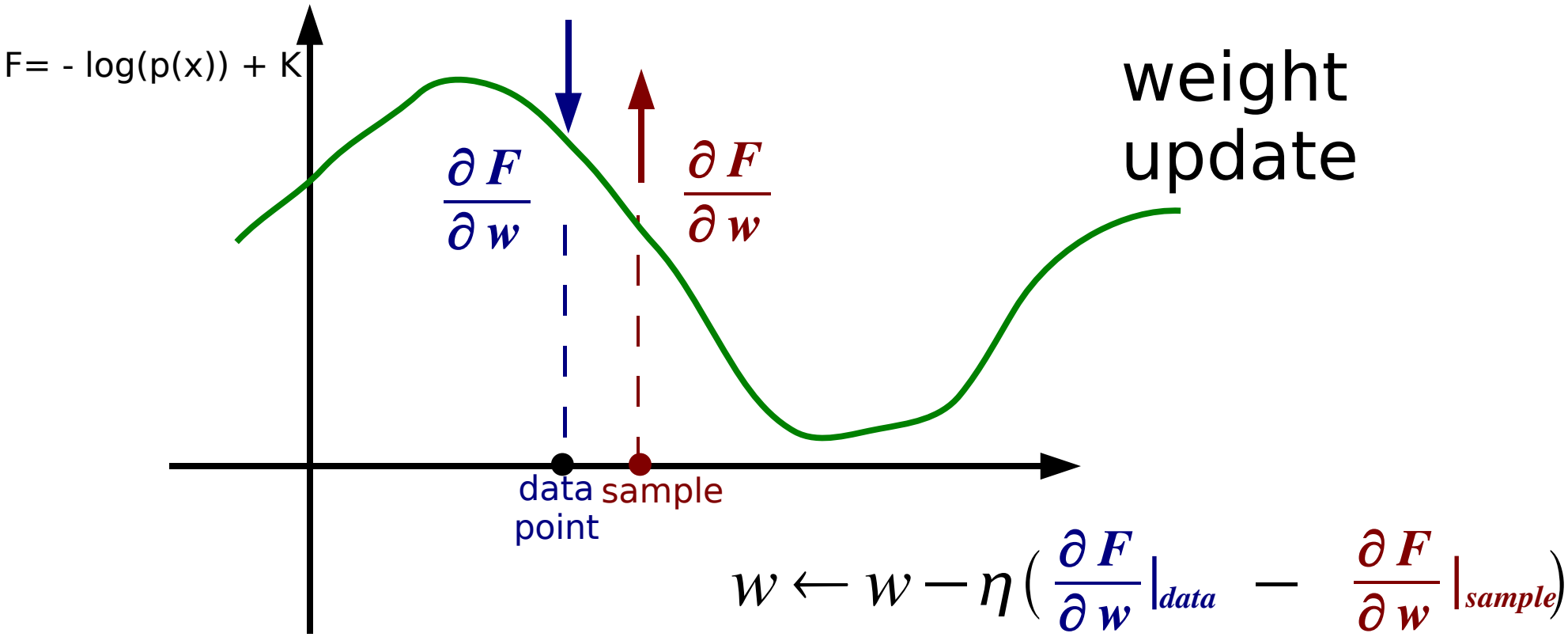
Learning


F= - log(p(x)) + K
$\frac{\partial F}{\partial w}$
HMC
dynamics
training sample

Learning

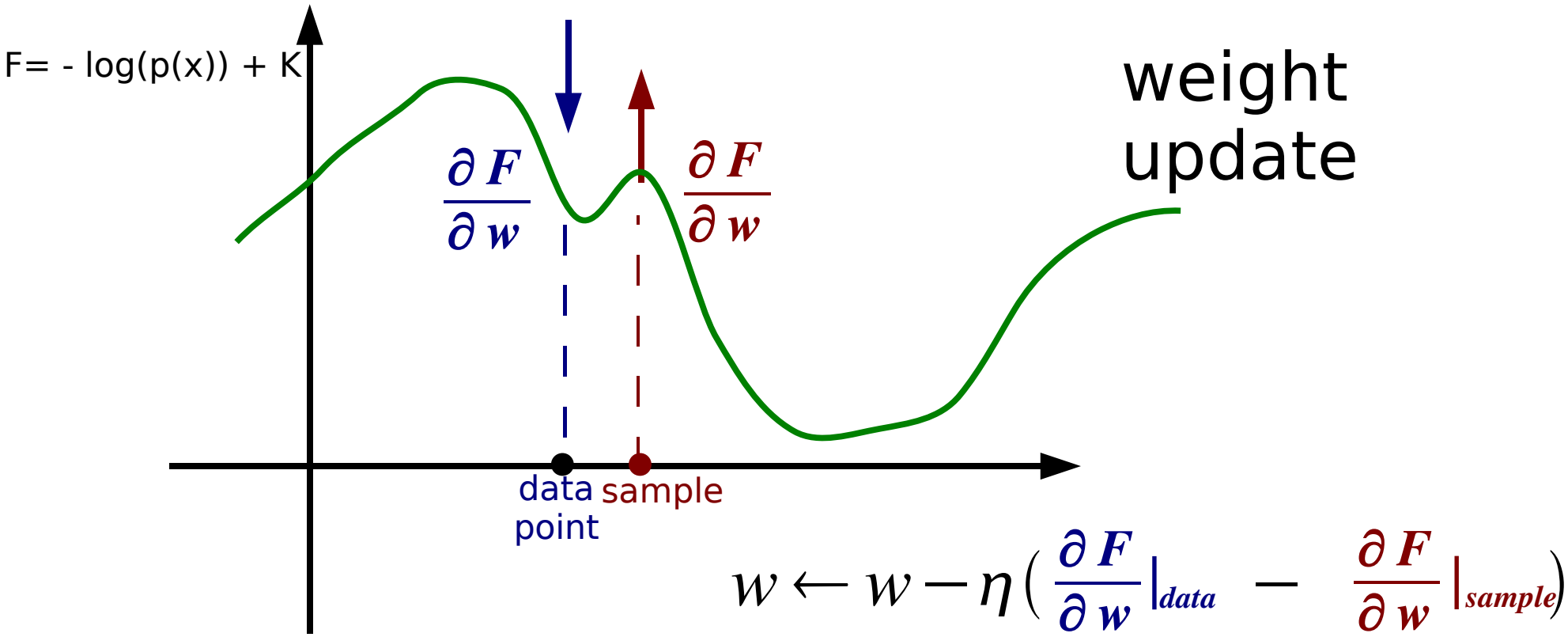


Learning



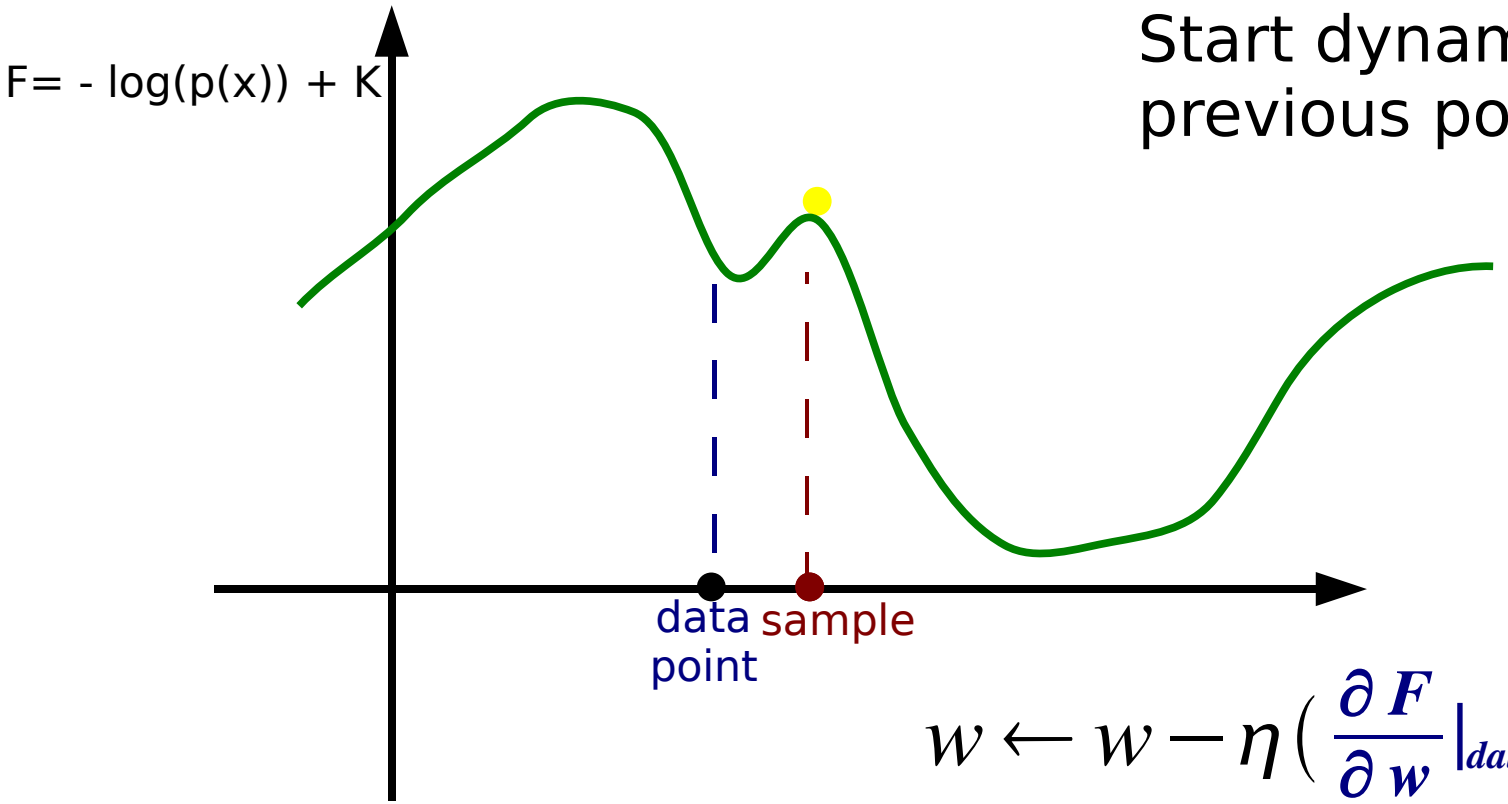
$$w \leftarrow w - \eta \left(\frac{\partial F}{\partial w}\Big|_{data} - \frac{\partial F}{\partial w}\Big|_{sample} \right)$$

Learning



$$w \leftarrow w - \eta \left(\frac{\partial F}{\partial w}\Big|_{data} - \frac{\partial F}{\partial w}\Big|_{sample} \right)$$

Learning



Start dynamics from previous point

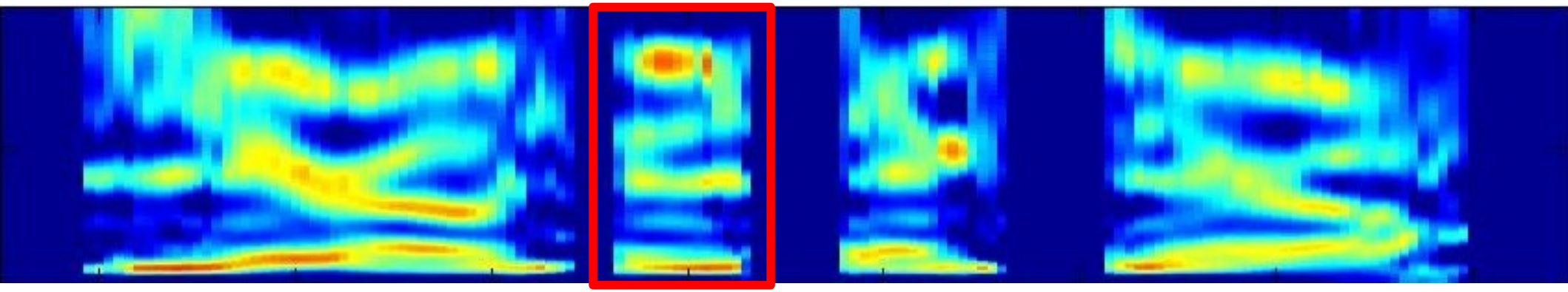
$$w \leftarrow w - \eta \left(\frac{\partial F}{\partial w}\Big|_{data} - \frac{\partial F}{\partial w}\Big|_{sample} \right)$$

Outline

- mathematical formulation of the model
- training
- learning acoustic features for spech recognition
- generation of natural images
- recognition of facial expression under occlusion
- conclusion

Speech Recognition on TIMIT

INPUT
- frame: 25ms Hamming window
- 10ms overlap between frames
- 39 log-filterbank outputs + energy per frame
- 15 frames
- PCA whitening → 384 components
- no deltas, no delta-deltas...

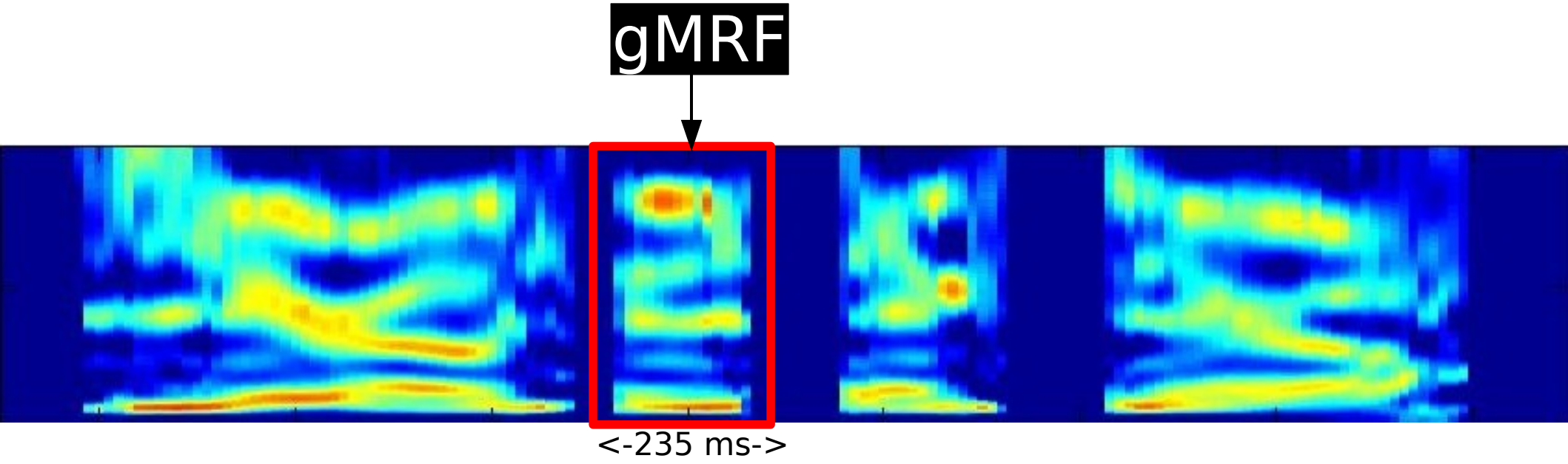


<-235 ms->

Dahl, Ranzato, Mohamed, Hinton, NIPS 2010

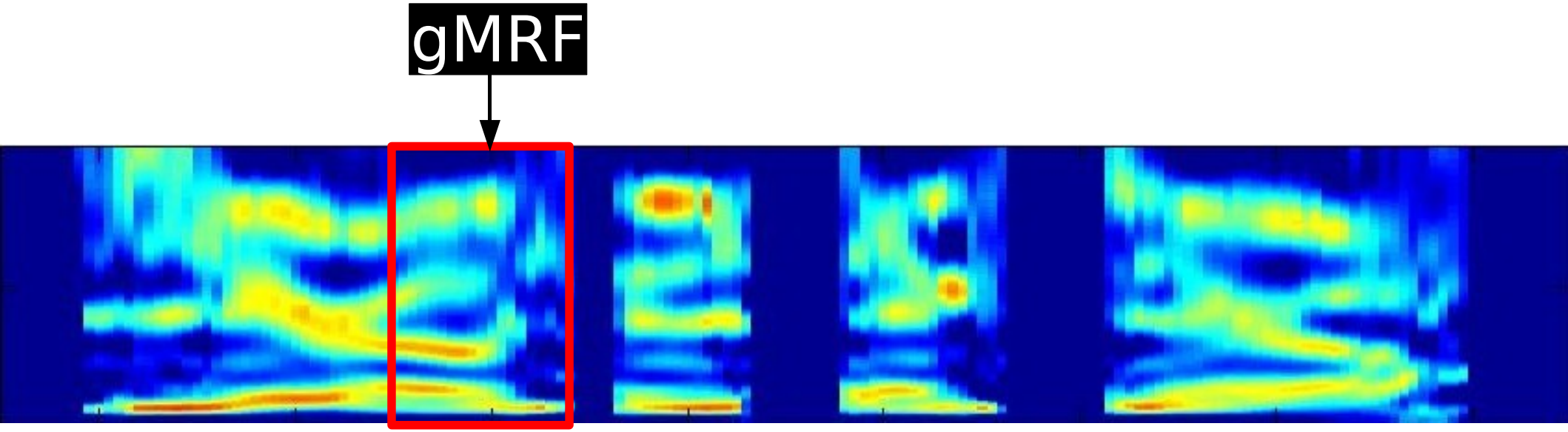
Speech Recognition on TIMIT

gated MRF with 1024 precision and 512 mean units
trained on independent windows



Speech Recognition on TIMIT

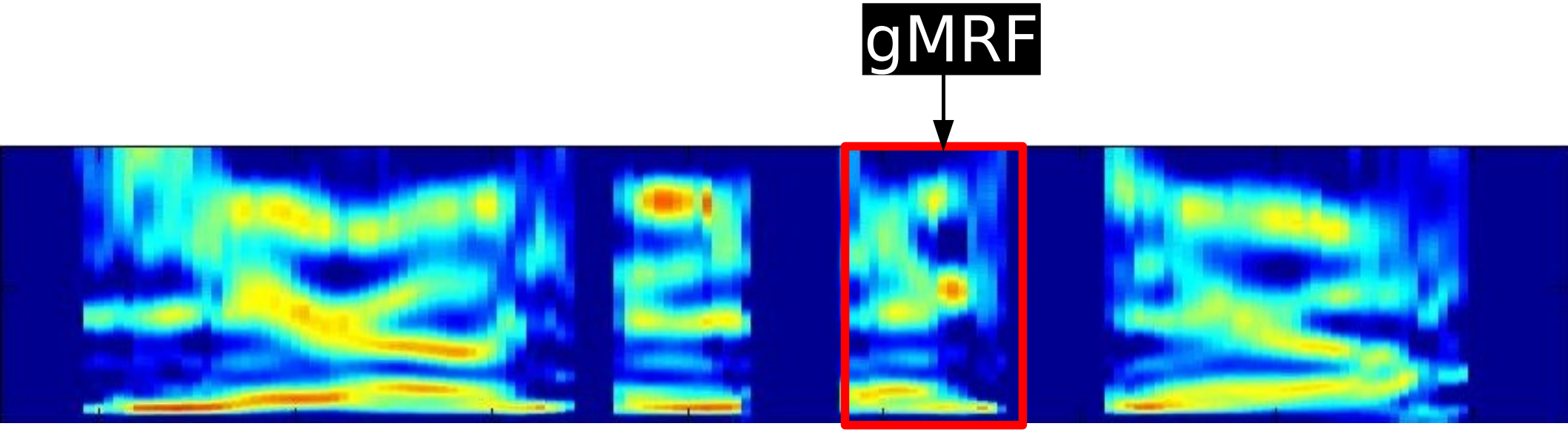
gated MRF with 1024 precision and 512 mean units
trained on independent windows



Dahl, Ranzato, Mohamed, Hinton, NIPS 2010

Speech Recognition on TIMIT

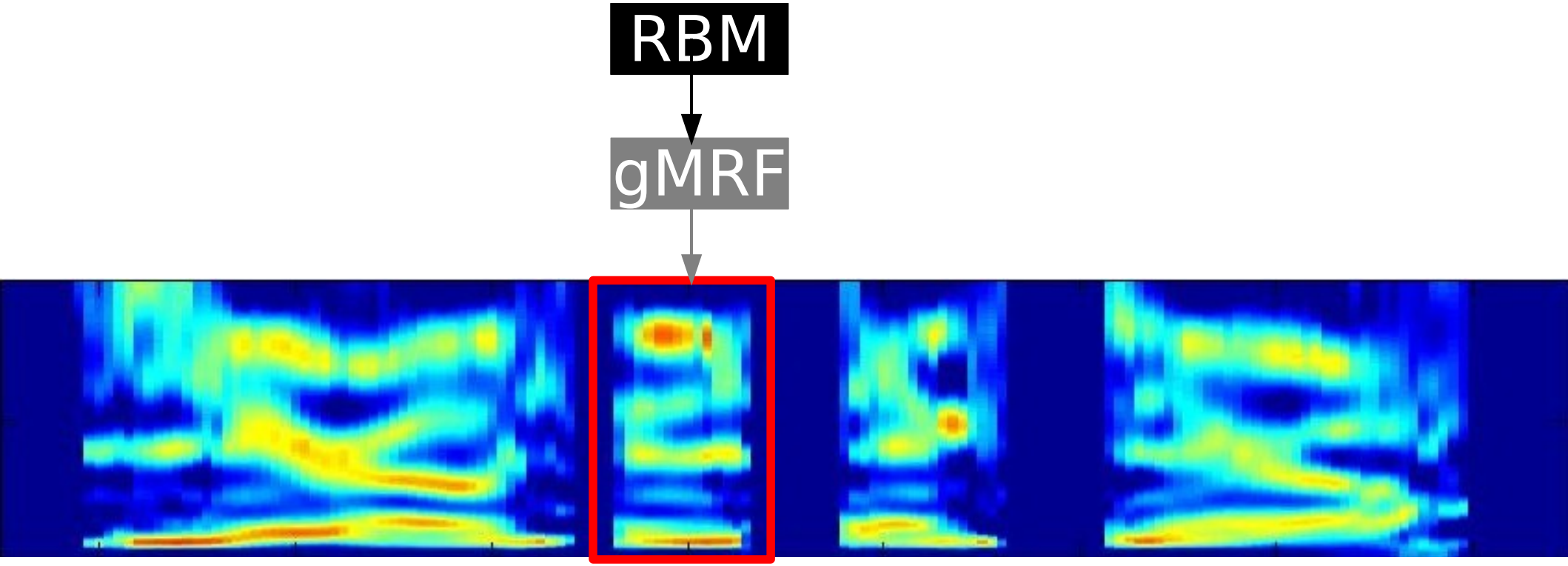
gated MRF with 1024 precision and 512 mean units
trained on independent windows



Dahl, Ranzato, Mohamed, Hinton, NIPS 2010

Speech Recognition on TIMIT

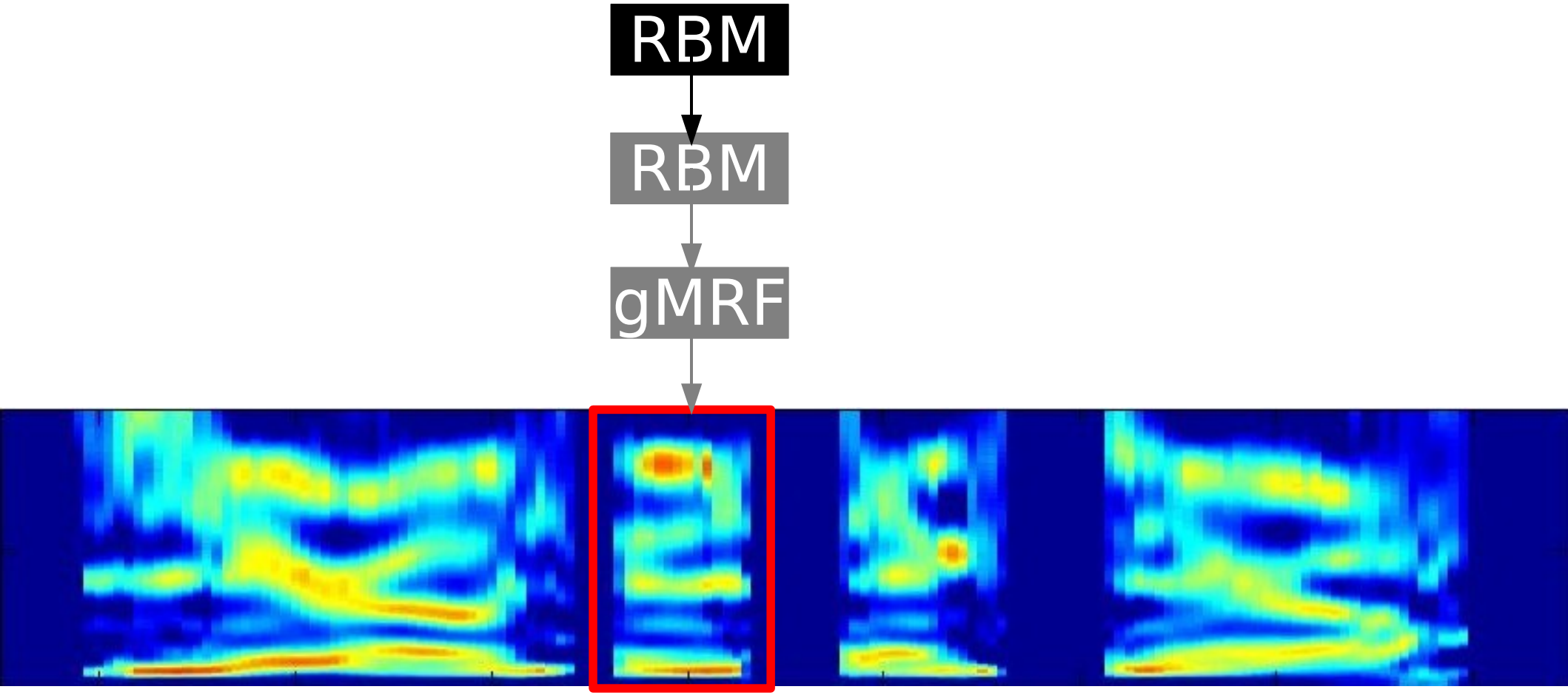
- For each training sample, infer latent variables of gated MRF
- Train 2nd layer binary-binary RBM (2048 units)



Dahl, Ranzato, Mohamed, Hinton, NIPS 2010

Speech Recognition on TIMIT

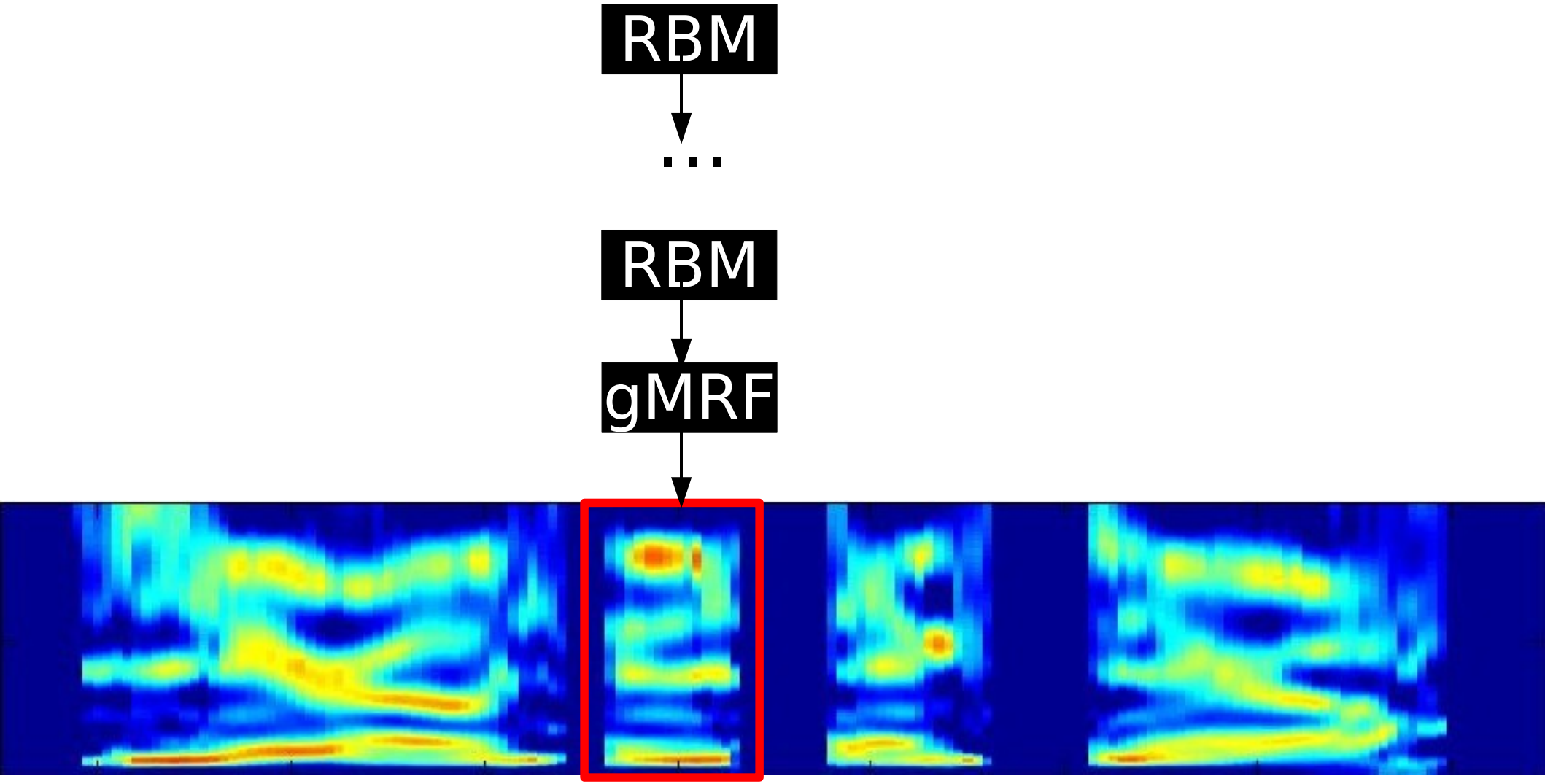
- Train 1st layer gated MRF
- Train 2nd layer binary-binary RBM (2048 units)
- Train 3rd layer binary-binary RBM (2048 units)



Dahl, Ranzato, Mohamed, Hinton, NIPS 2010

Speech Recognition on TIMIT

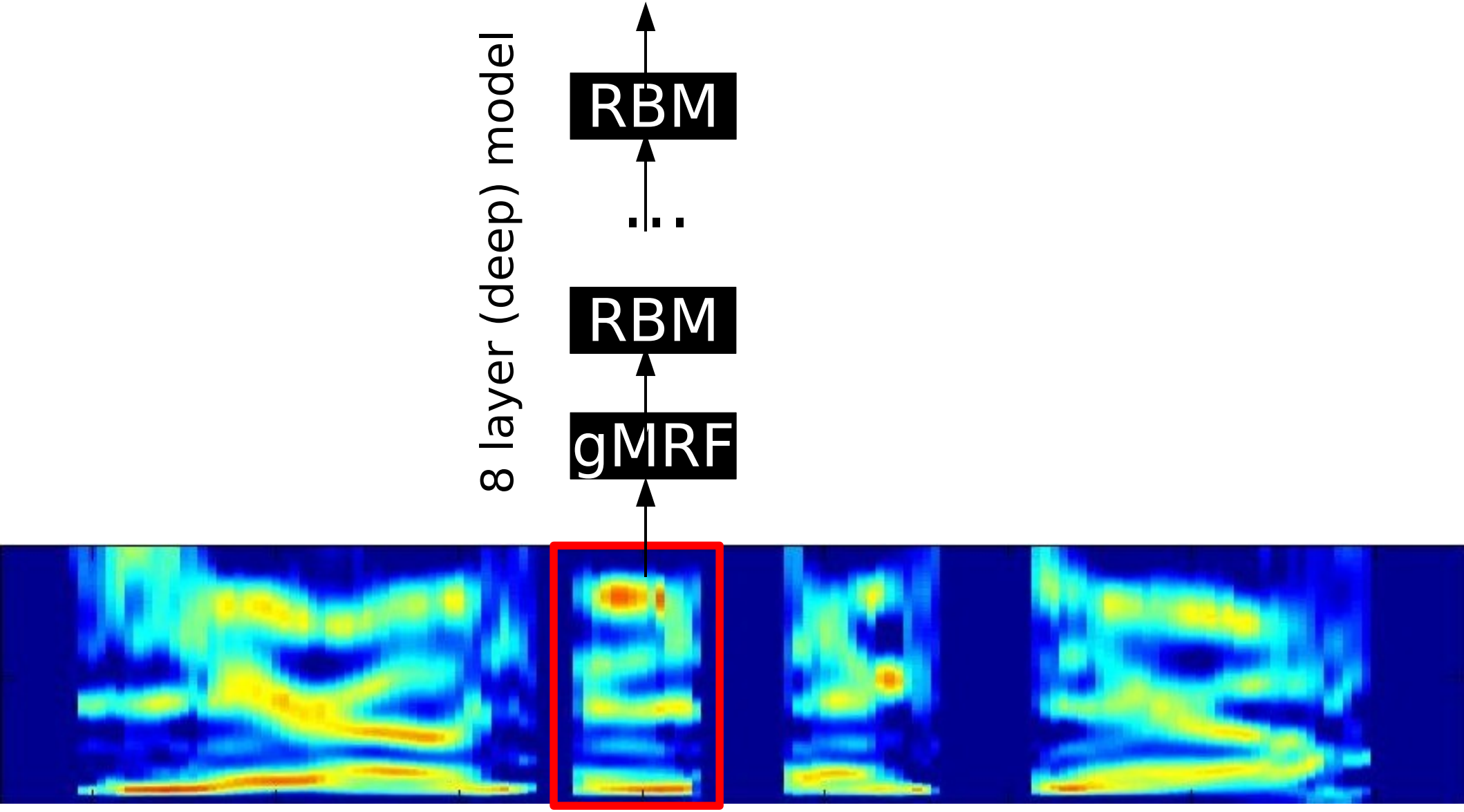
8 layer (deep) model



Dahl, Ranzato, Mohamed, Hinton, NIPS 2010

Speech Recognition on TIMIT

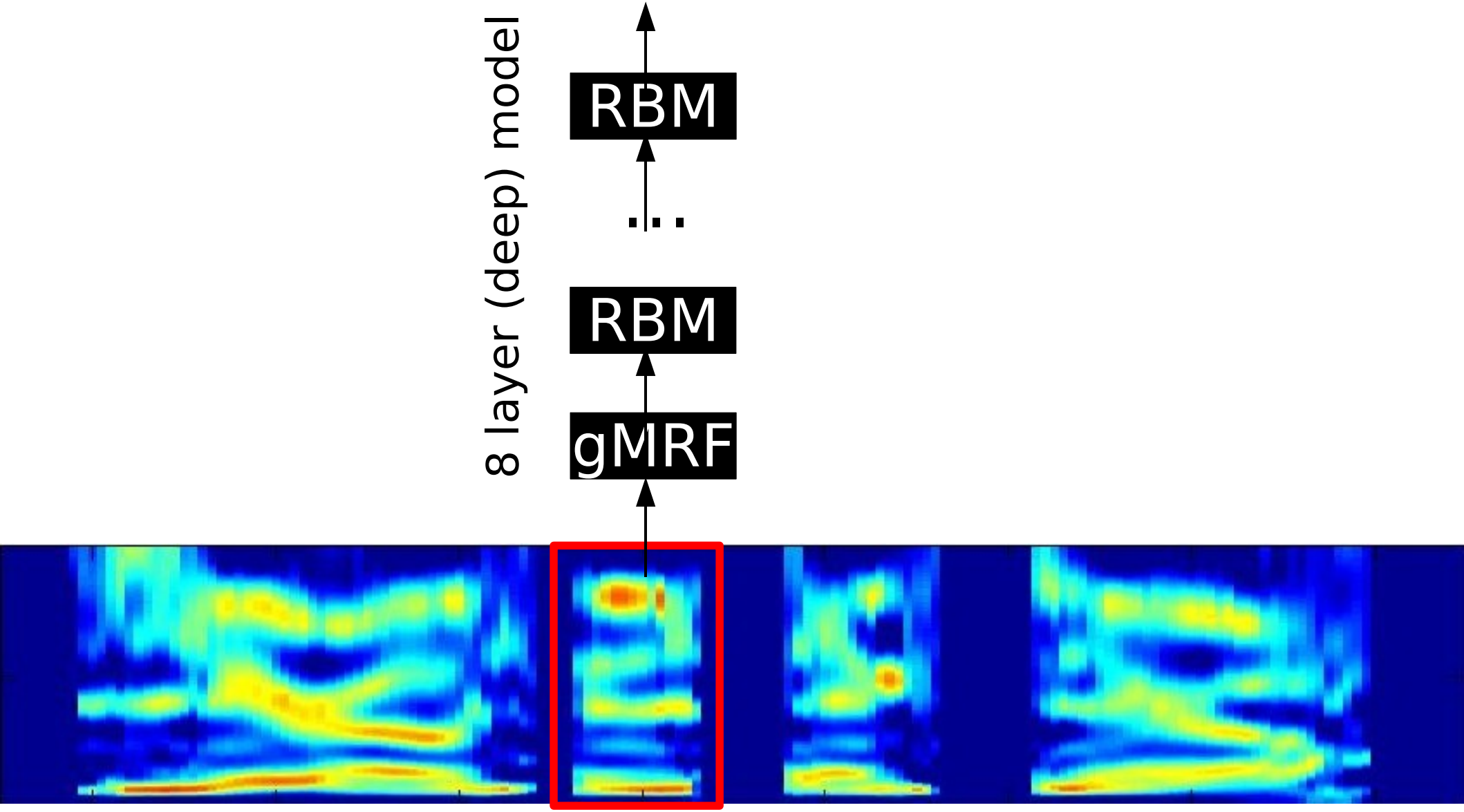
Supervised training: predict states of aligned HMM



Dahl, Ranzato, Mohamed, Hinton, NIPS 2010

Speech Recognition on TIMIT

Test: predict states of HMM → decoding (bigram language model)



Dahl, Ranzato, Mohamed, Hinton, NIPS 2010

Speech Recognition on TIMIT

METHOD	PER
CRF	34.8%
Large-Margin GMM	33.0%
CD-HMM	27.3%
Augmented CRF	26.6%
RNN	26.1%
Bayesian Triphone HMM	25.6%
Triphone HMM discrim. trained	22.7%
DBN with gated MRF	20.5%

Speech Recognition on TIMIT

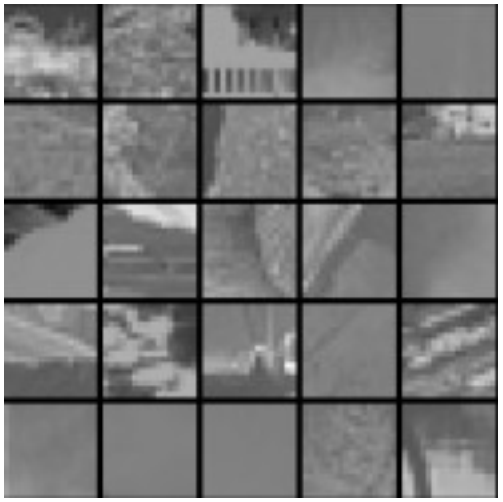
METHOD	PER	Year
CRF	34.8%	2008
Large-Margin GMM	33.0%	2006
CD-HMM	27.3%	2009
Augmented CRF	26.6%	2009
RNN	26.1%	1994
Bayesian Triphone HMM	25.6%	1998
Triphone HMM discrim. trained	22.7%	2009
DBN with gated MRF	20.5%	2010

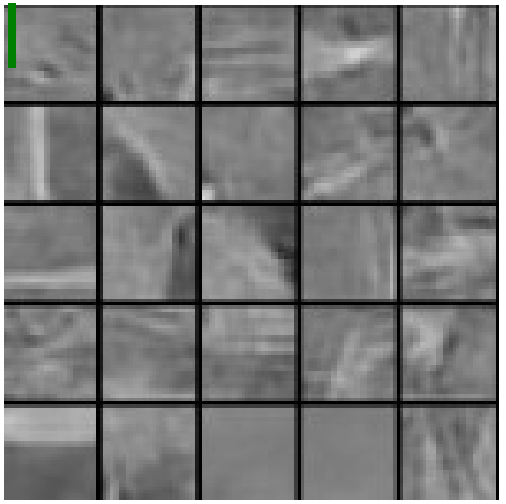
Outline

- mathematical formulation of the model
- training
- learning acoustic features for spech recognition
- generation of natural images
- recognition of facial expression under occlusion
- conclusion

Generation natural image patches

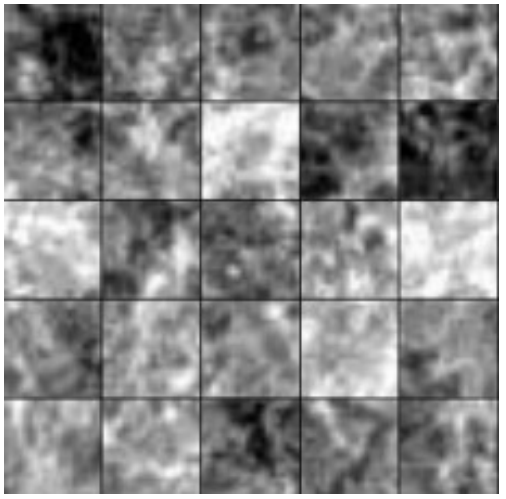
Natural images





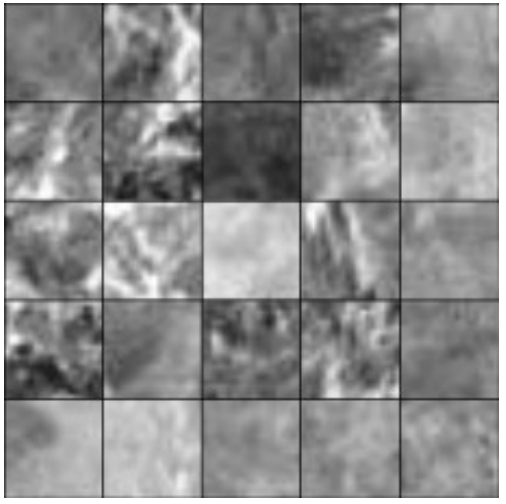
mcRBM

Ranzato and Hinton CVPR 2010



GRBM

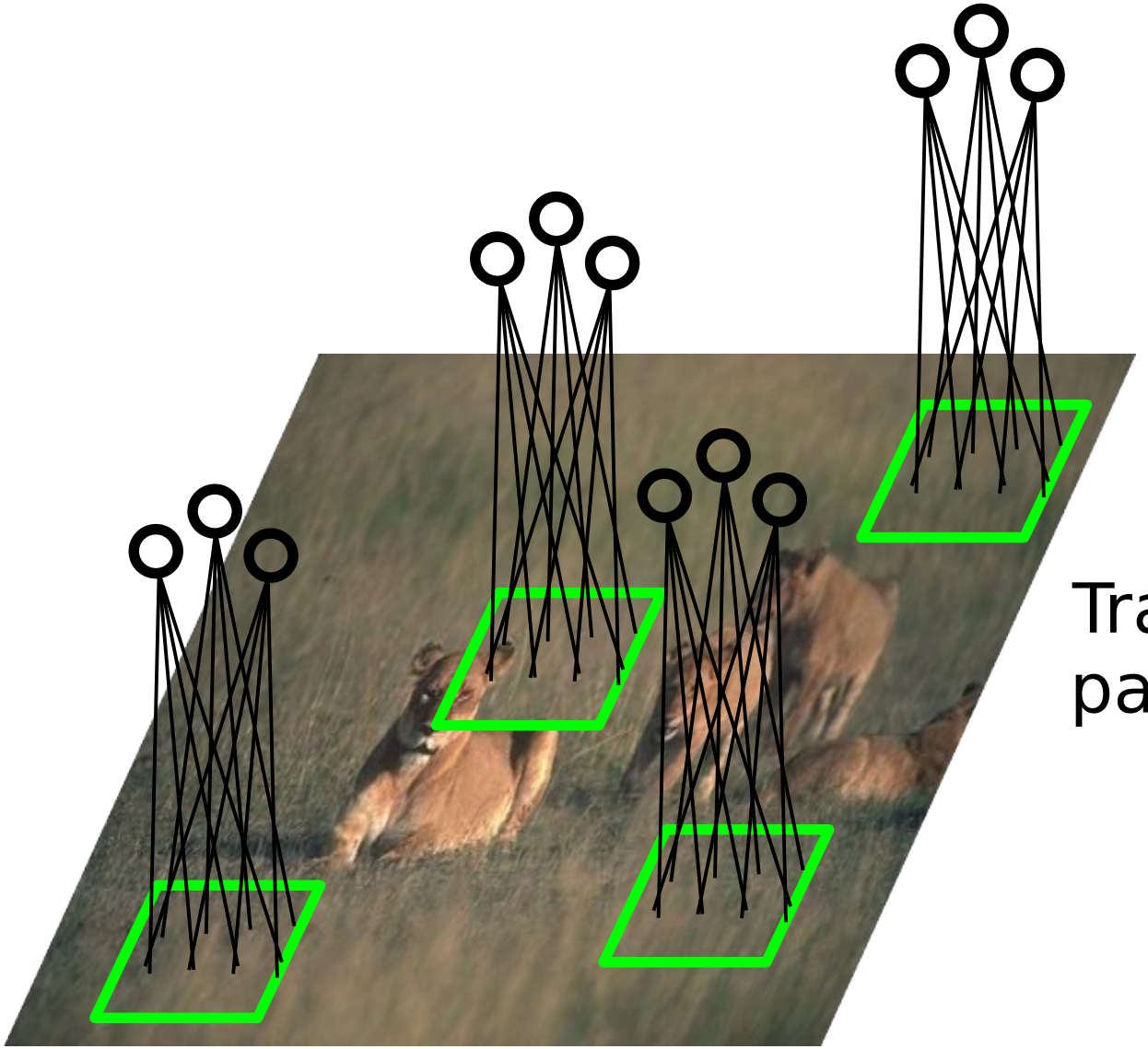
from Osindero and Hinton NIPS 2008



S-RBM + DBN

from Osindero and Hinton NIPS 2008

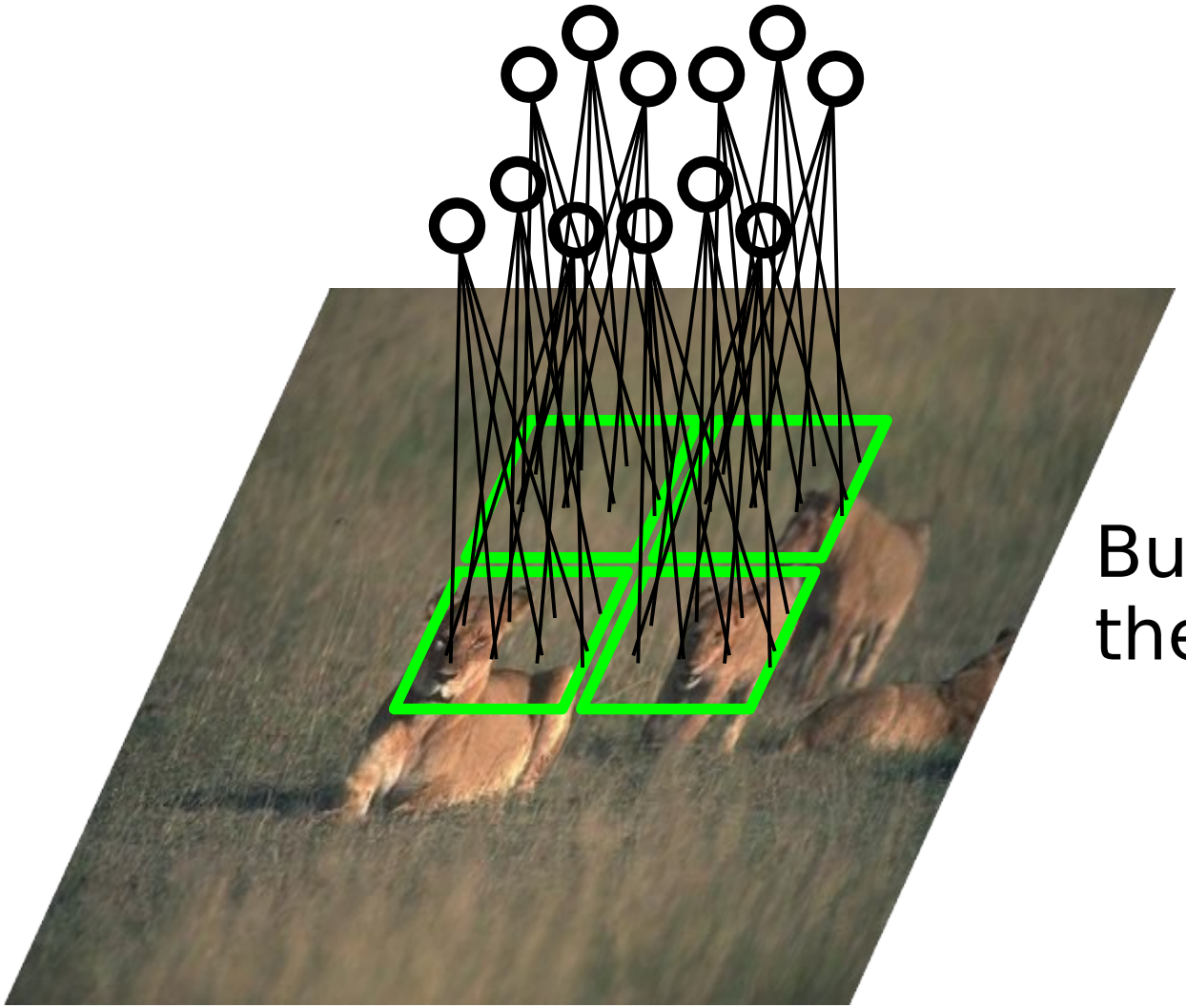
From Patches to High-Resolution Images



Training by picking
patches at random

From Patches to High-Resolution Images

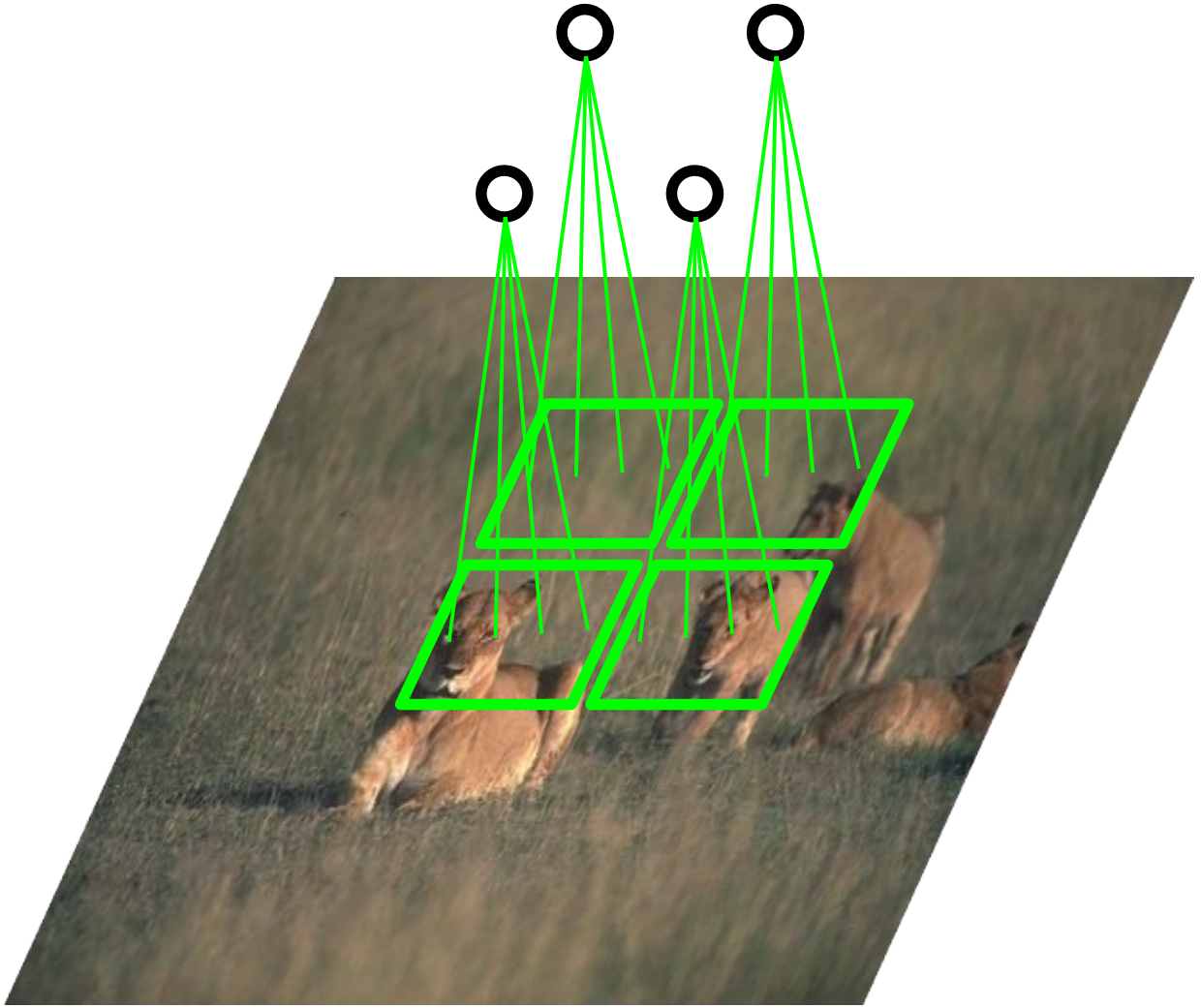
This is not a good way to extend the model to big images: block artifacts



But we could also take them from a grid

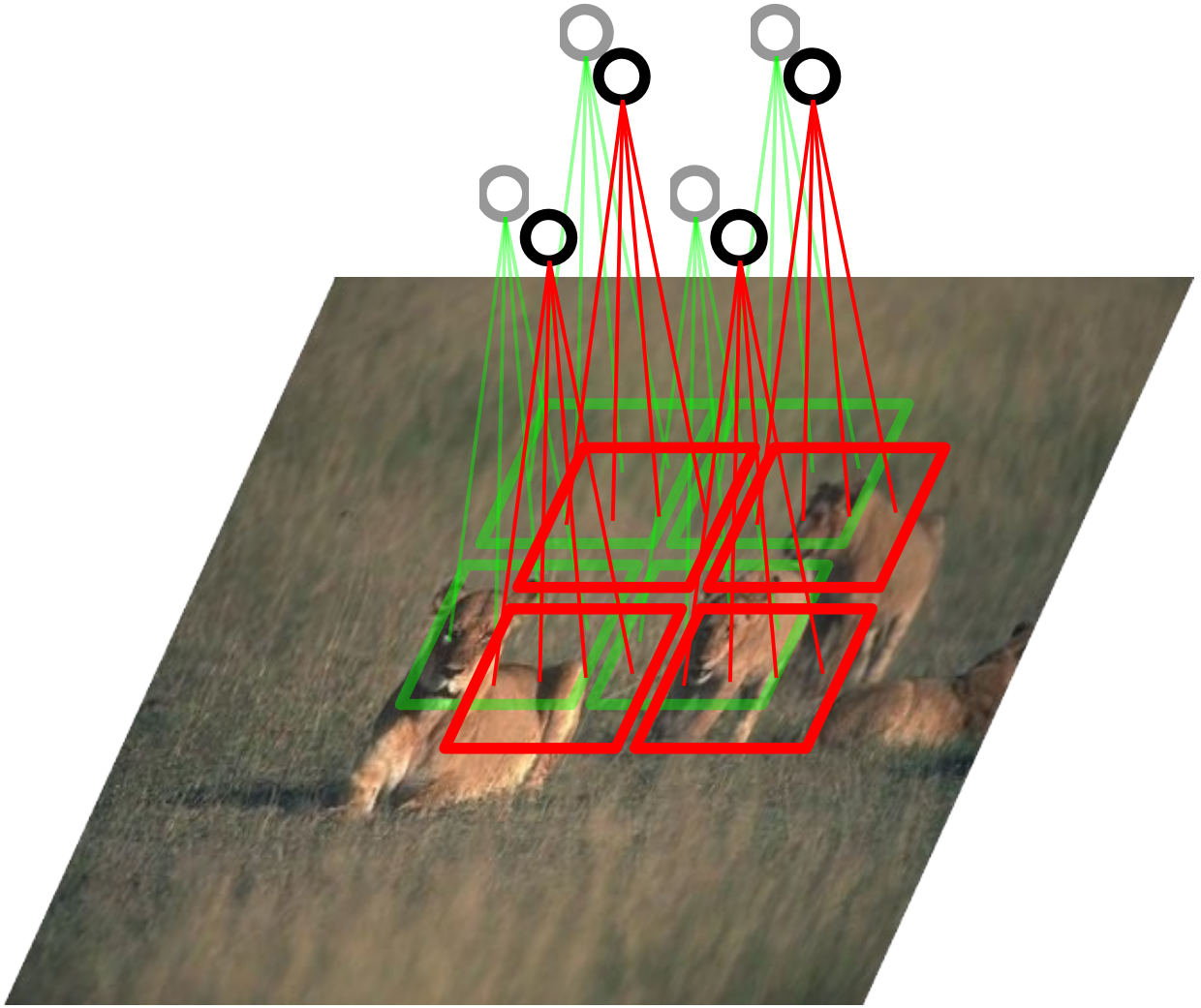
From Patches to High-Resolution Images

IDEA: have one subset of filters applied to these locations,



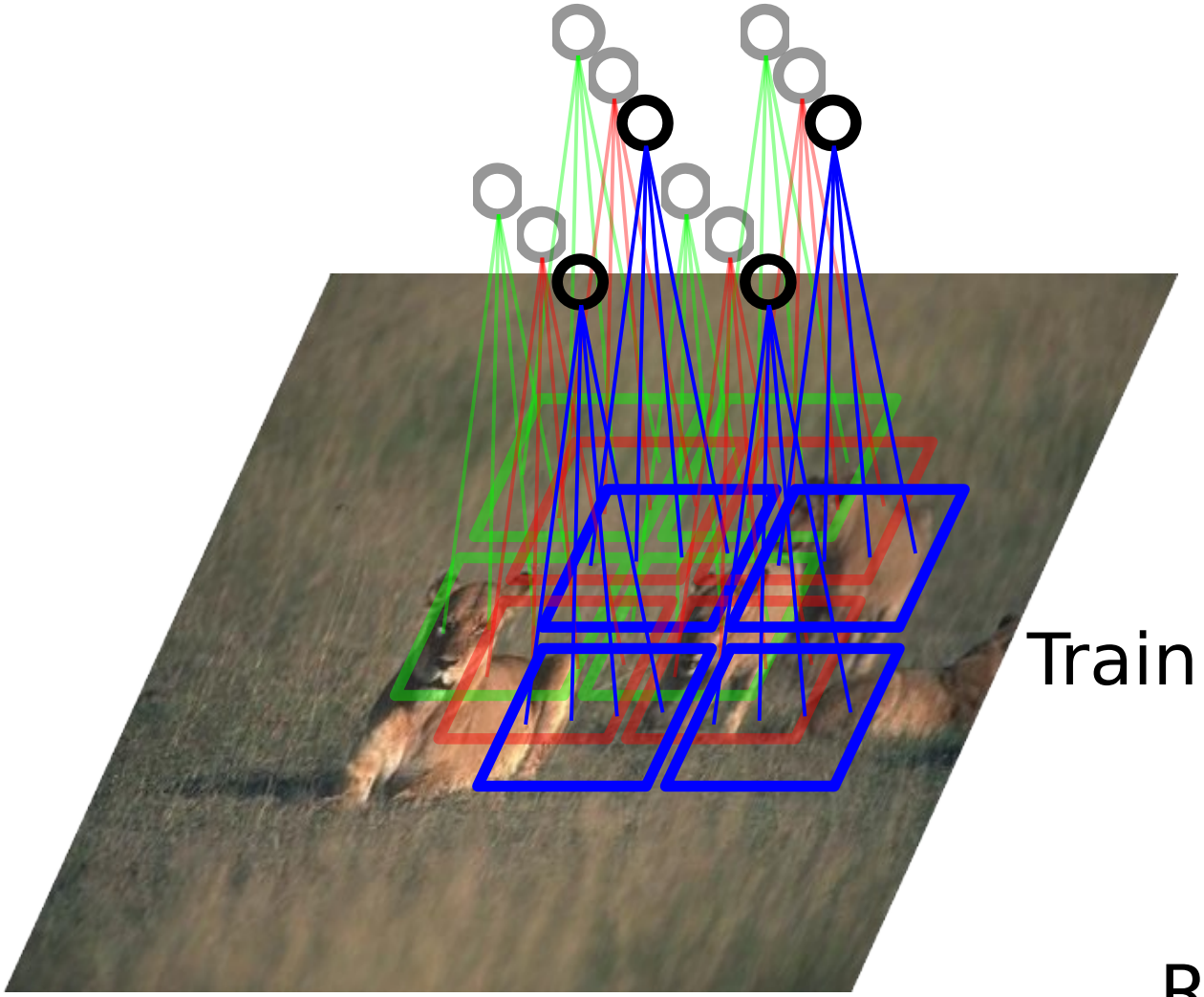
From Patches to High-Resolution Images

IDEA: have one subset of filters applied to these locations, another subset to these locations



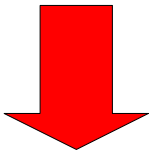
From Patches to High-Resolution Images

IDEA: have one subset of filters applied to these locations, another subset to these locations, etc.



Gregor LeCun arXiv 2010
Ranzato, Mnih, Hinton NIPS 2010

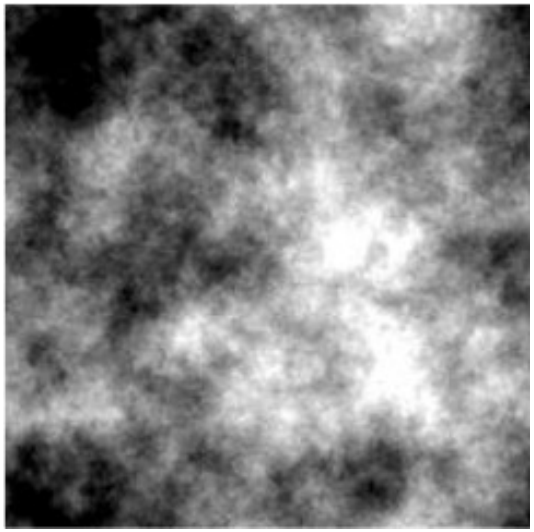
Train jointly all parameters.



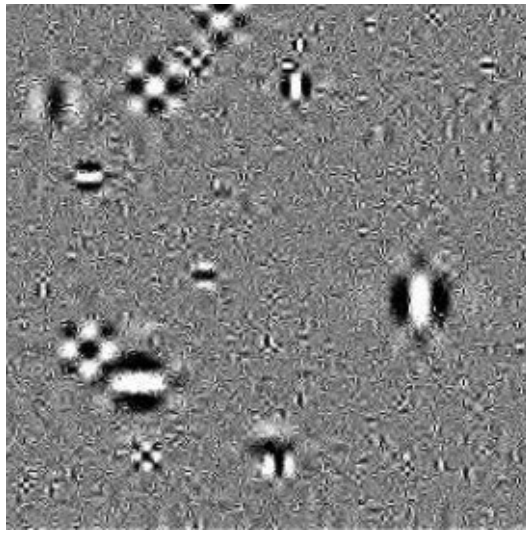
No block artifacts
Reduced redundancy

Sampling High-Resolution Images

Gaussian model



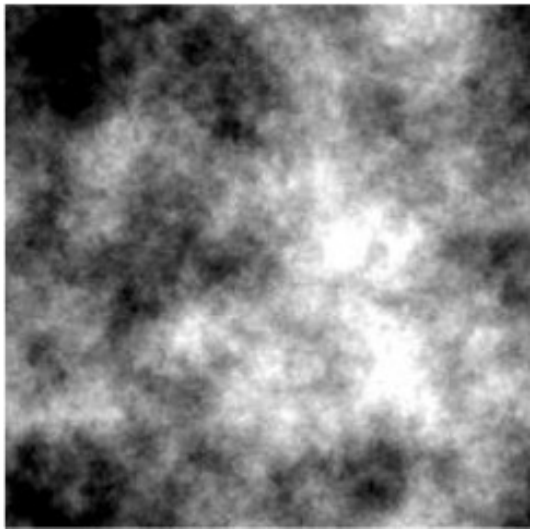
marginal wavelet



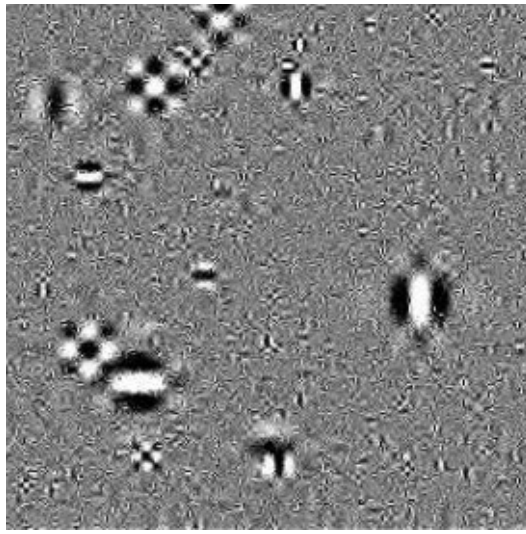
from Simoncelli 2005

Sampling High-Resolution Images

Gaussian model

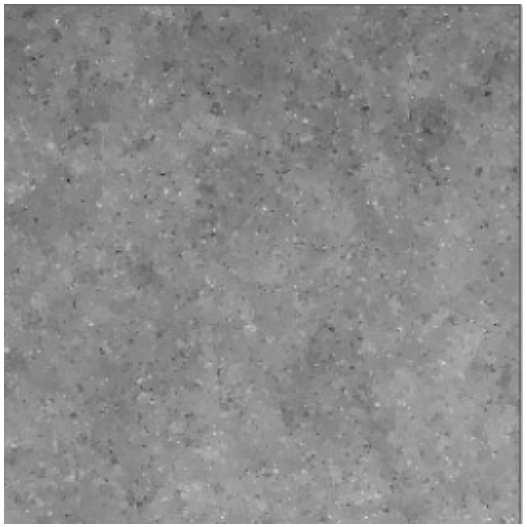


marginal wavelet



from Simoncelli 2005

Pair-wise MRF



FoE

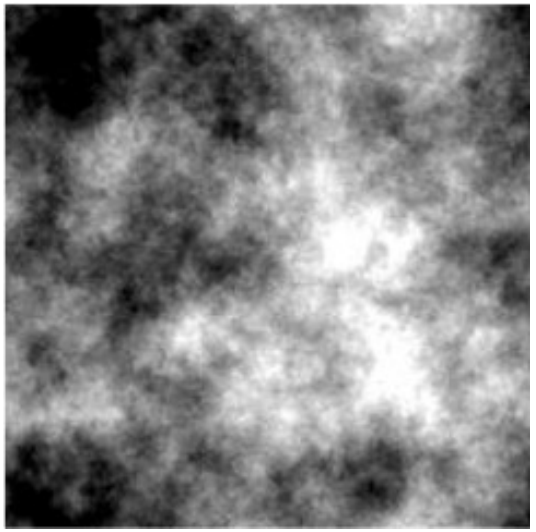
from Schmidt, Gao, Roth CVPR 2010

Sampling High-Resolution Images

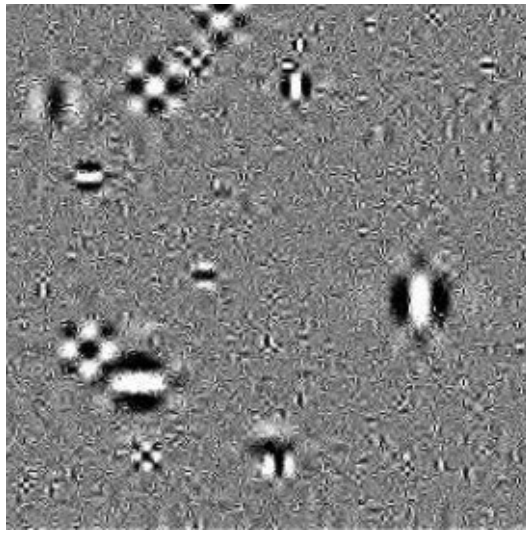
Mean Covariance Model

Ranzato, Mnih, Hinton NIPS 2010

Gaussian model

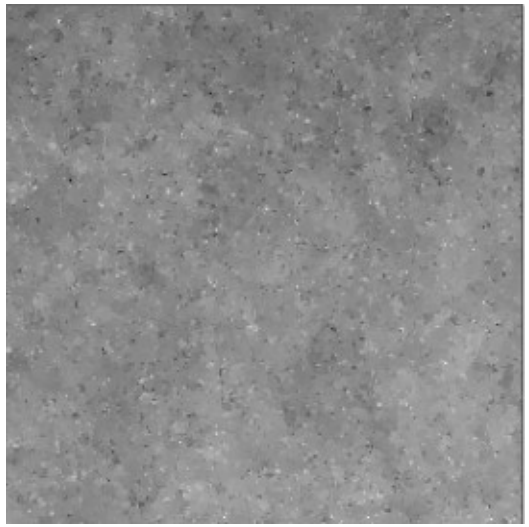


marginal wavelet



from Simoncelli 2005

Pair-wise MRF

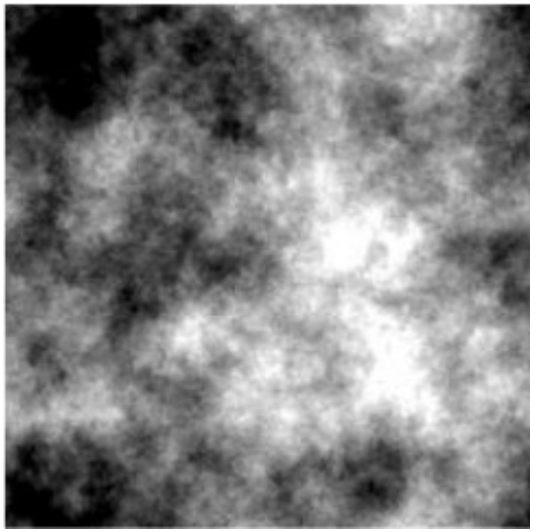


FoE

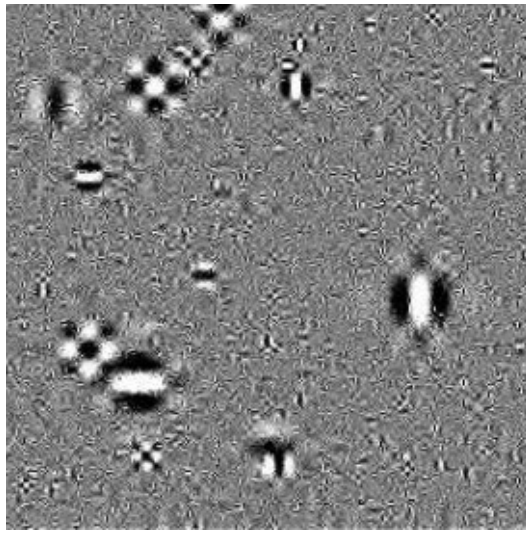
from Schmidt, Gao, Roth CVPR 2010

Sampling High-Resolution Images

Gaussian model



marginal wavelet

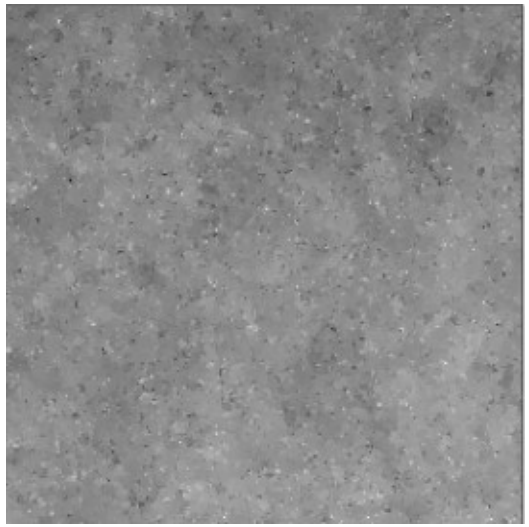


from Simoncelli 2005

Mean Covariance Model

Ranzato, Mnih, Hinton NIPS 2010

Pair-wise MRF

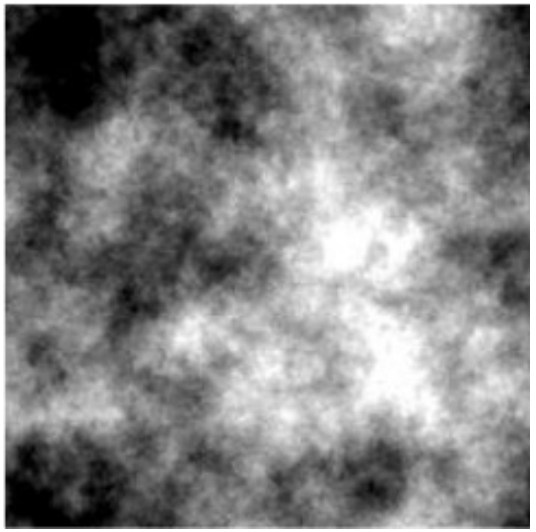


FoE

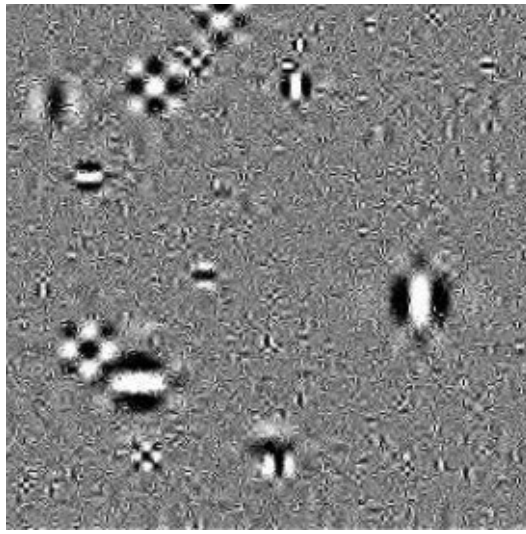
from Schmidt, Gao, Roth CVPR 2010

Sampling High-Resolution Images

Gaussian model

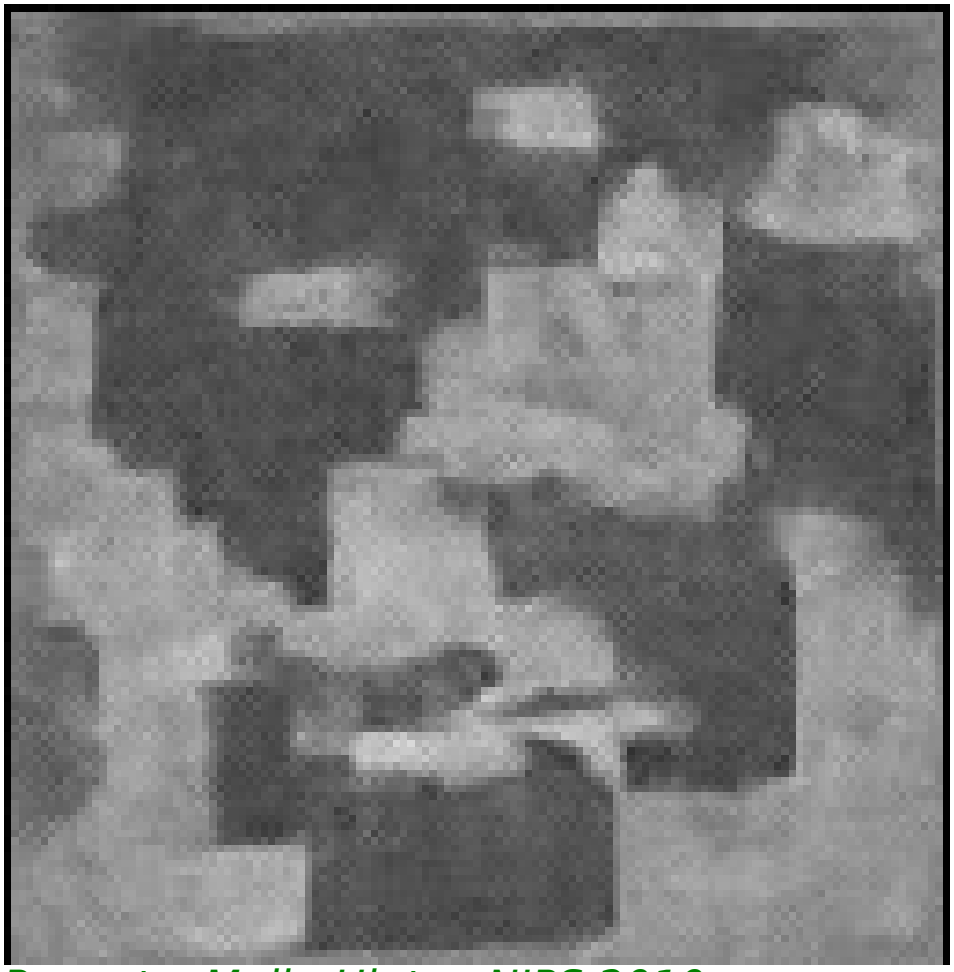


marginal wavelet



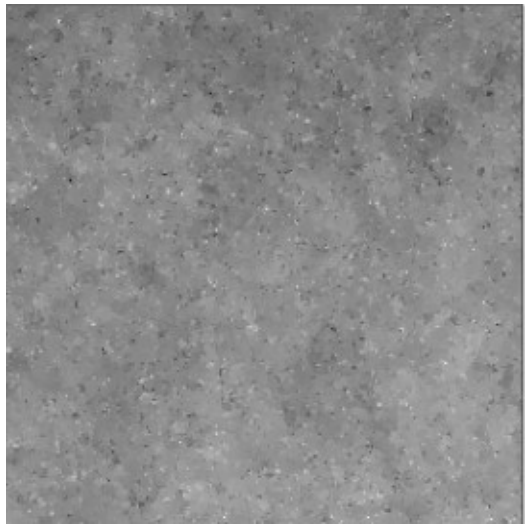
from Simoncelli 2005

Mean Covariance Model



Ranzato, Mnih, Hinton NIPS 2010

Pair-wise MRF

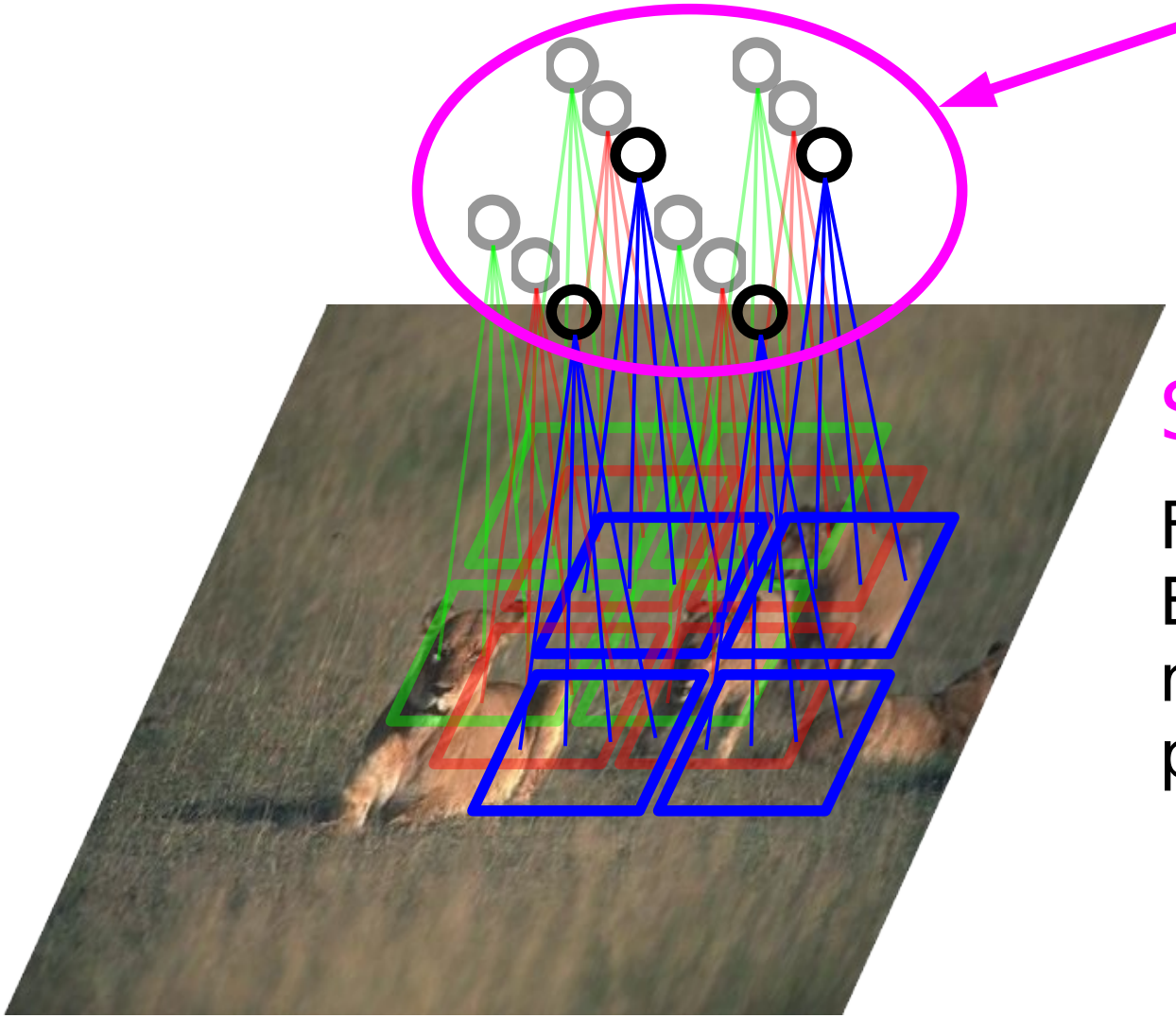


FoE

from Schmidt, Gao, Roth CVPR 2010

Making the model.. “DEEPER “

Treat these units as data
to train a similar model on the top

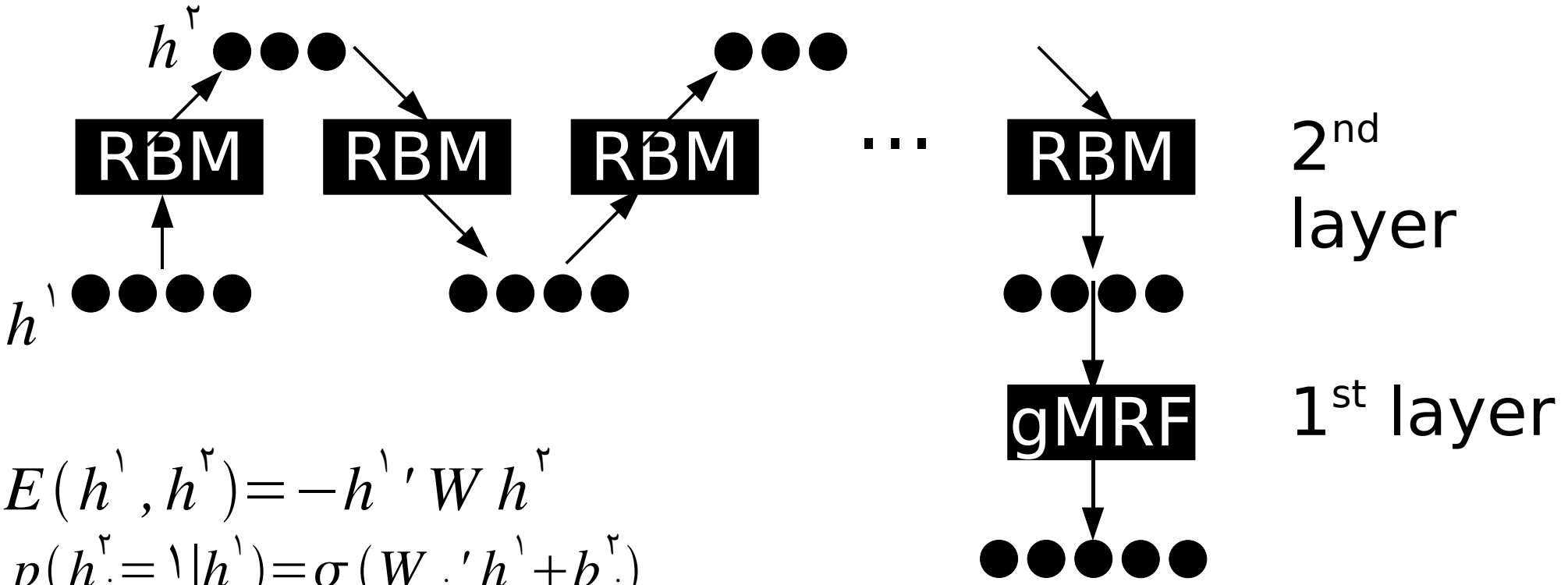


SECOND STAGE

Field of binary RBM's.
Each hidden unit has a receptive field of 30x30 pixels in input space.

Sampling from the DEEPER model

- Sample from 2nd layer
- project sample in image space using 1st layer p(x|h)



$$E(h^{1}, h^{2}) = -h^{1\prime} W h^{2}$$

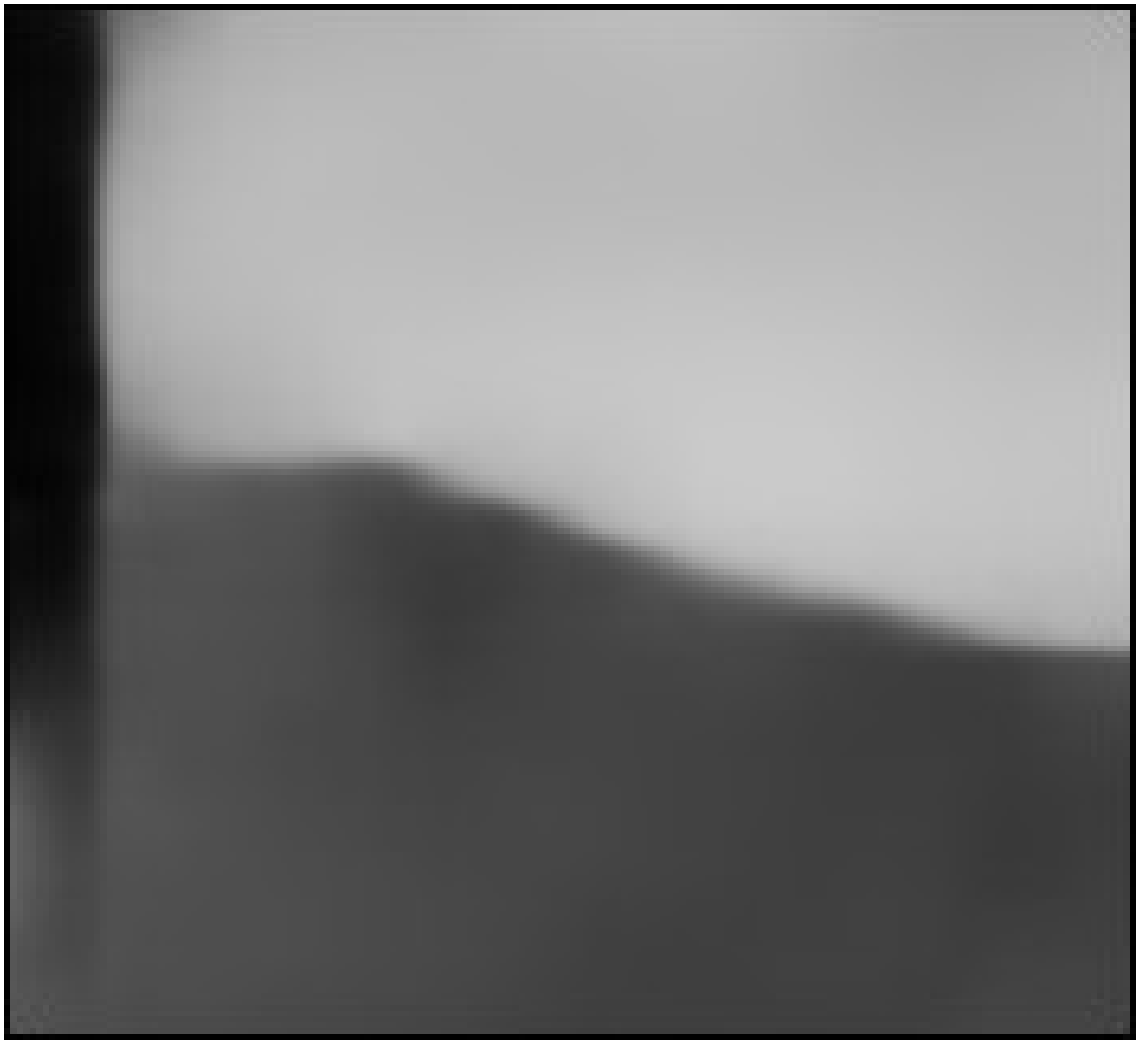
$$p(h^{2}_{j} = 1 | h^{1}) = \sigma(W_{j}' h^{1} + b^{2}_{j})$$

$$p(h^{1}_{k} = 1 | h^{2}) = \sigma(W_{k} h^{2} + b^{1}_{k})$$

Samples from Deep Generative Model

1st stage model

3rd stage model



Samples from Deep Generative Model

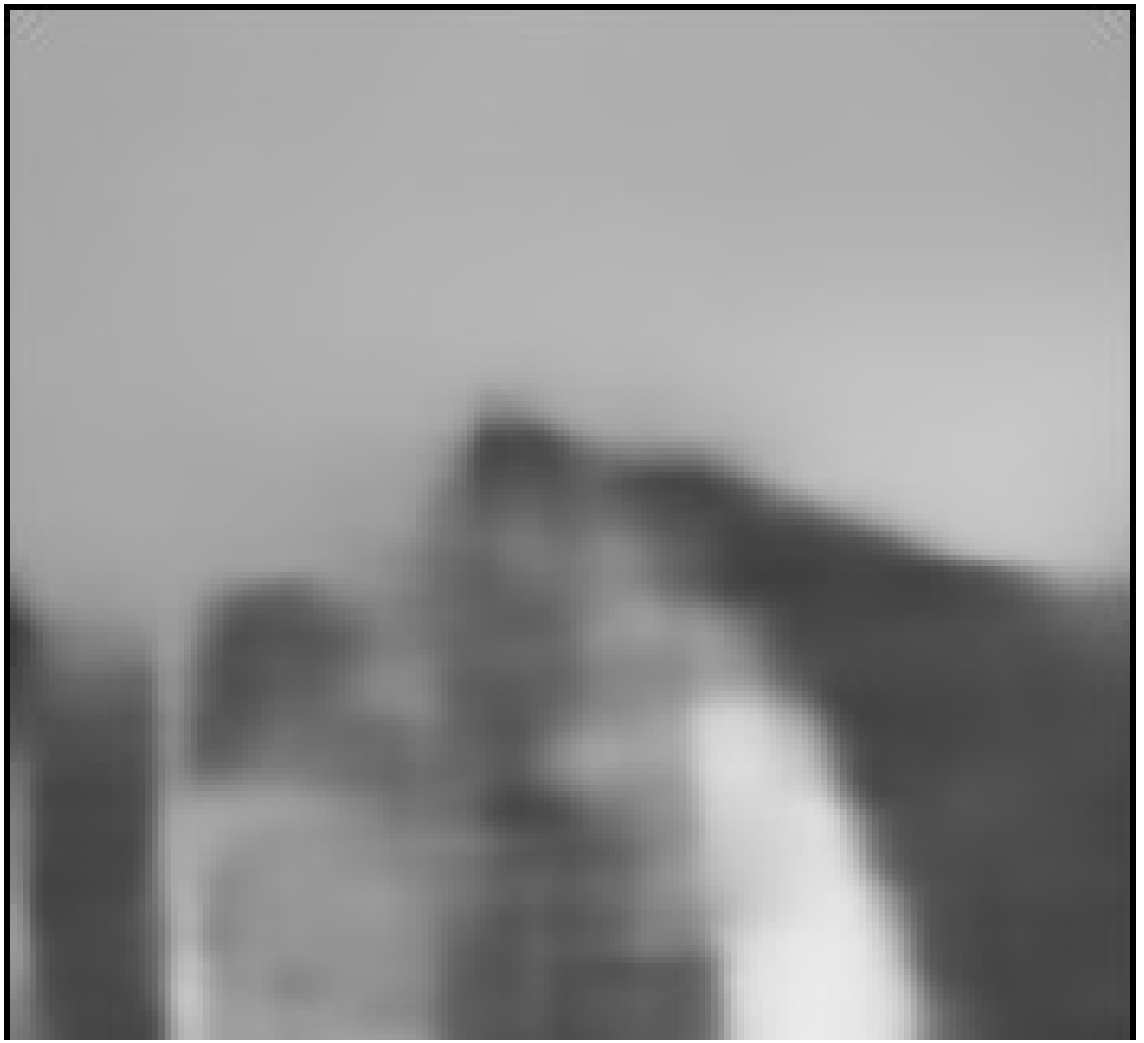
1st stage model

3rd stage model

Samples from Deep Generative Model

1st stage model

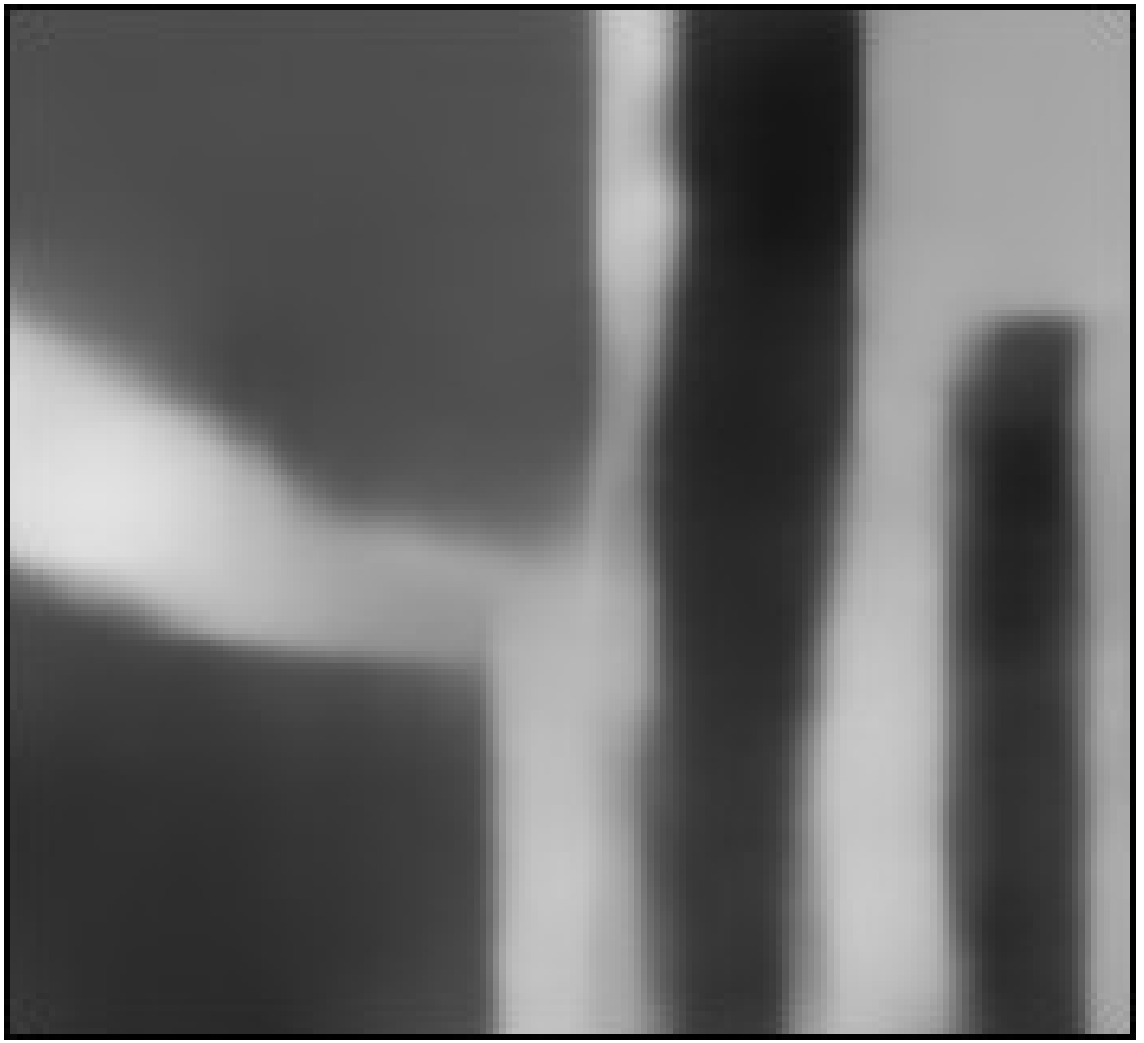
3rd stage model



Samples from Deep Generative Model

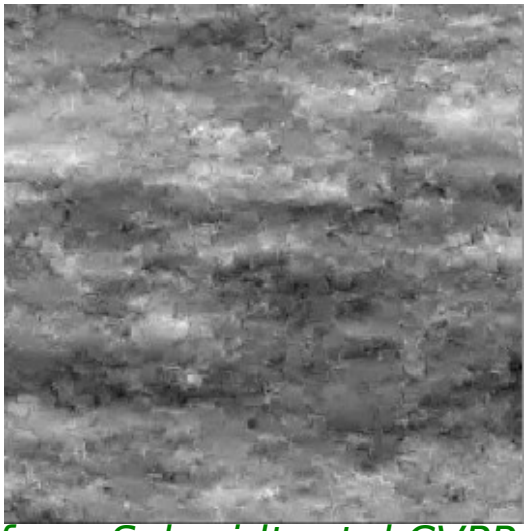
1st stage model

3rd stage model



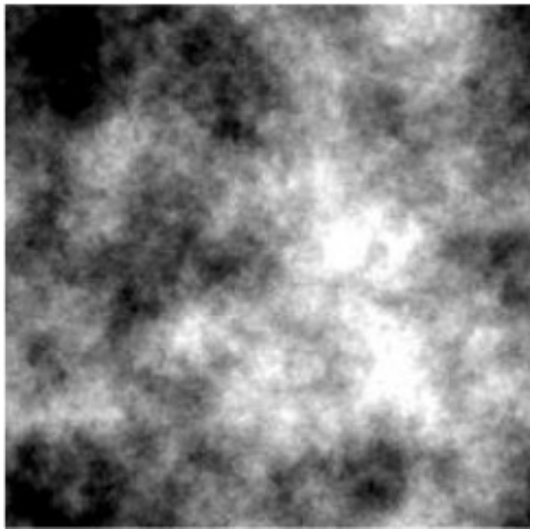
Sampling High-Resolution Images

FoE



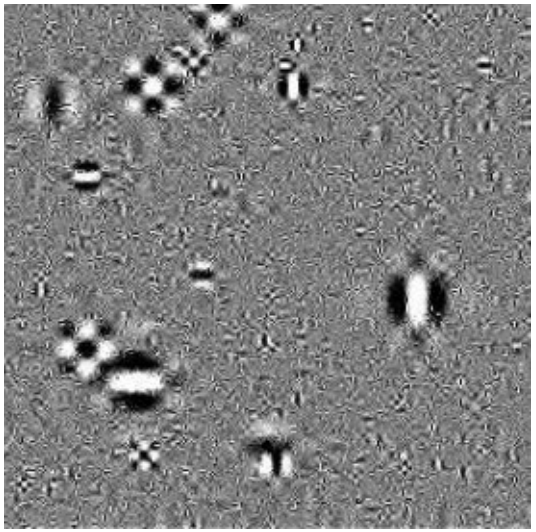
from Schmidt, etal CVPR 2010

Gaussian model



from Simoncelli 2005

marginal wavelet

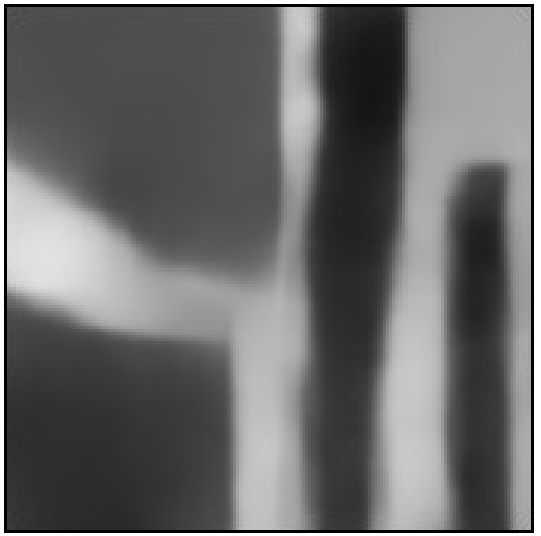


from Simoncelli 2005

Deep - 1

Ranzato, Mnih, Hinton NIPS 2010

Deep - 3 layers



Ranzato, et al. CVPR 2011

Outline

- mathematical formulation of the model
- training
- learning acoustic features for spech recognition
- generation of natural images
- recognition of facial expression under occlusion
- conclusion

Facial Expression Recognition

Toronto Face Dataset (J. Susskind et al. 2010)

~ 100K unlabeled faces from different sources

~ 4K labeled images

Resolution: 48x48 pixels

7 facial expressions

anger

Ranzato, et al. CVPR 2011

Facial Expression Recognition

Toronto Face Dataset (J. Susskind et al. 2010)

~ 100K unlabeled faces from different sources

~ 4K labeled images

Resolution: 48x48 pixels

7 facial expressions

disgust

Facial Expression Recognition

Toronto Face Dataset (J. Susskind et al. 2010)

~ 100K unlabeled faces from different sources

~ 4K labeled images

Resolution: 48x48 pixels

7 facial expressions

fear

Facial Expression Recognition

Toronto Face Dataset (J. Susskind et al. 2010)

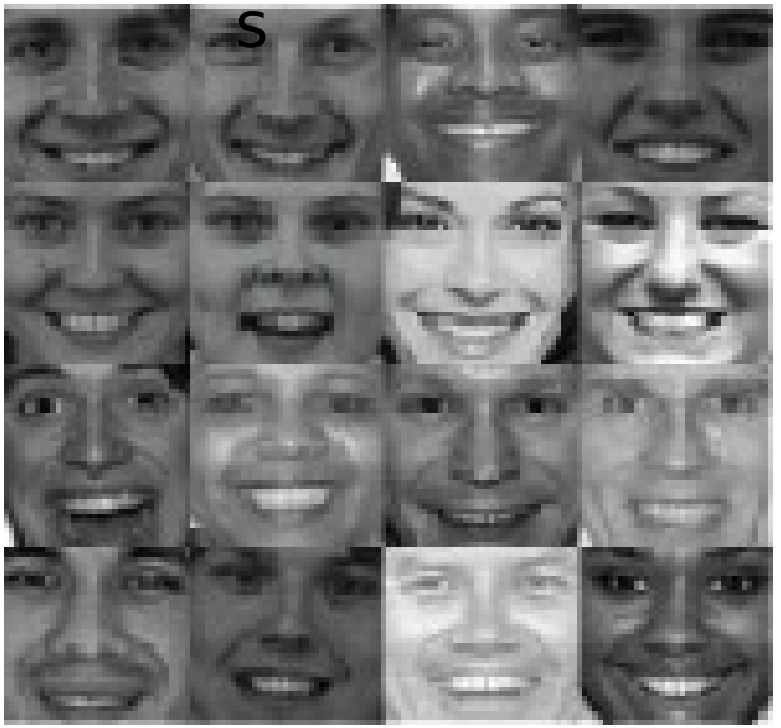
~ 100K unlabeled faces from different sources

~ 4K labeled images

Resolution: 48x48 pixels

7 facial expressions

happines
s



Facial Expression Recognition

Toronto Face Dataset (J. Susskind et al. 2010)

~ 100K unlabeled faces from different sources

~ 4K labeled images

Resolution: 48x48 pixels

7 facial expressions

neutral

Facial Expression Recognition

Toronto Face Dataset (J. Susskind et al. 2010)

~ 100K unlabeled faces from different sources

~ 4K labeled images

Resolution: 48x48 pixels

7 facial expressions

sadness

Facial Expression Recognition

Toronto Face Dataset (J. Susskind et al. 2010)

~ 100K unlabeled faces from different sources

~ 4K labeled images

Resolution: 48x48 pixels

7 facial expressions

surprise

Facial Expression Recognition

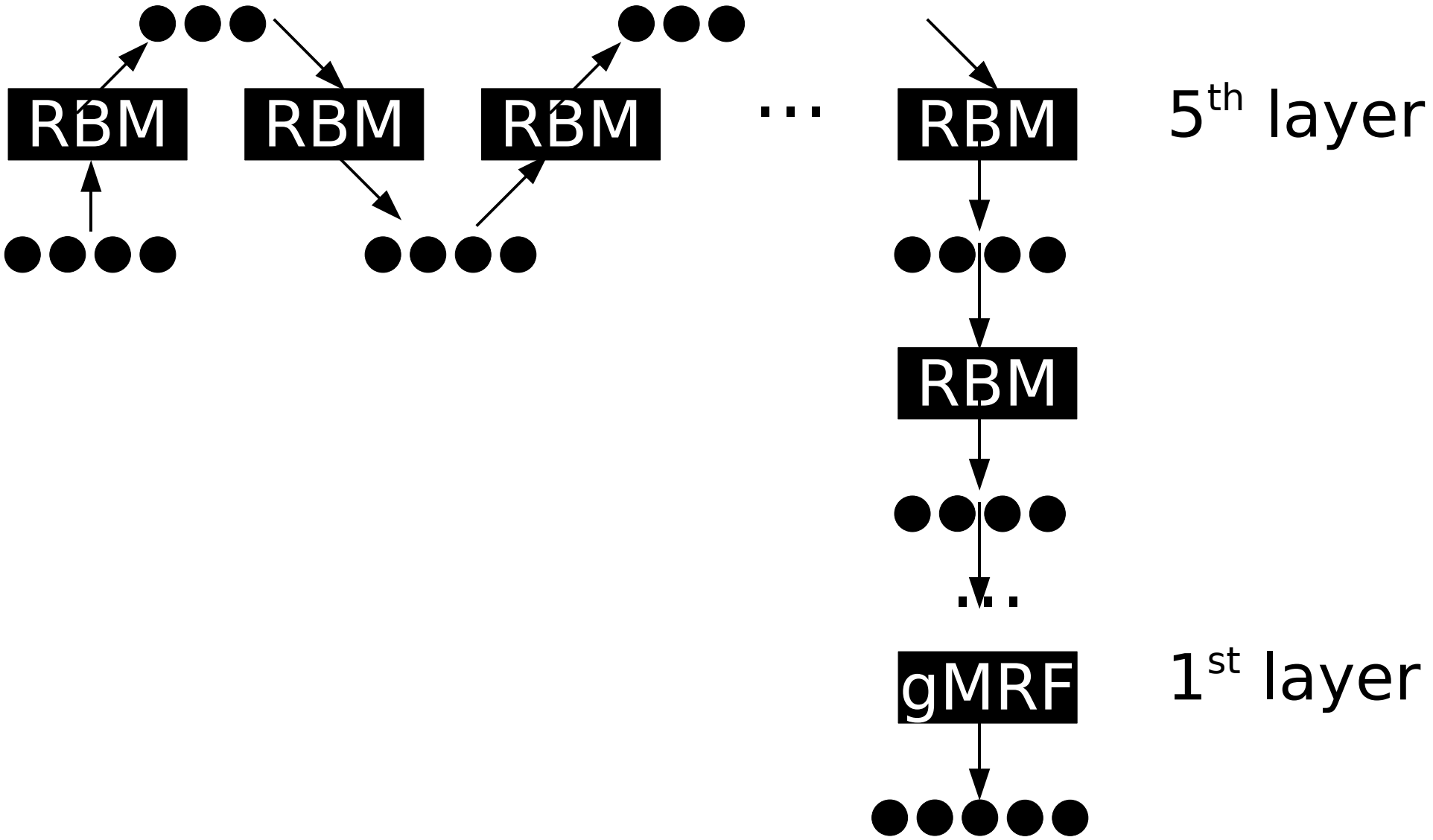
- 1st layer using local (not shared) connectivity
- layers above are fully connected
- 5 layers in total

- Result

- Linear Classifier on raw pixels	71.5%
- Gaussian RBF SVM on raw pixels	76.2%
- Gabor + PCA + linear classifier	80.1%
Dailey et al. J. Cog. Science 2002	
- Sparse coding	74.6%
Wright et al. PAMI 2008	
- DEEP model (3 layers):	82.5%

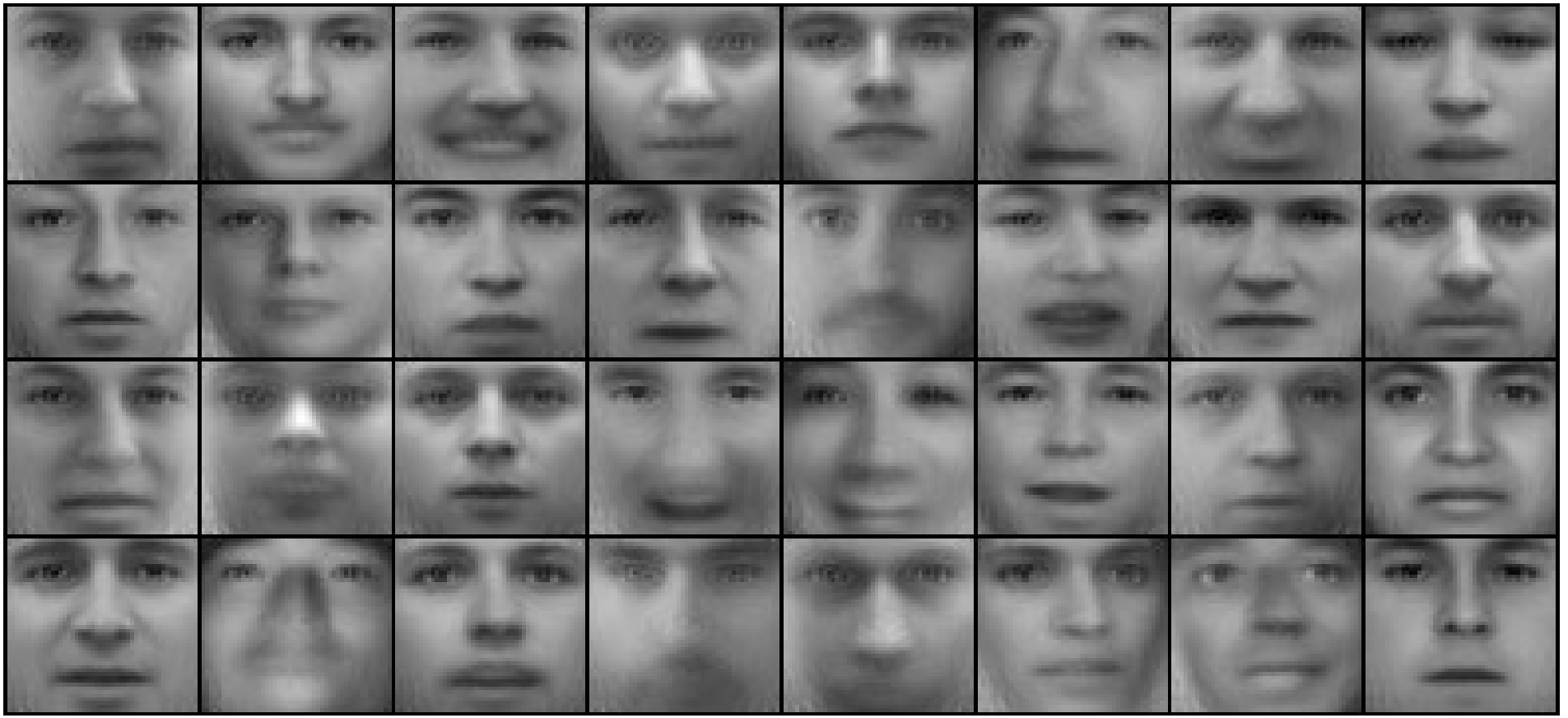
Facial Expression Recognition

- We can draw samples from the model (5th layer with 128 hiddens)



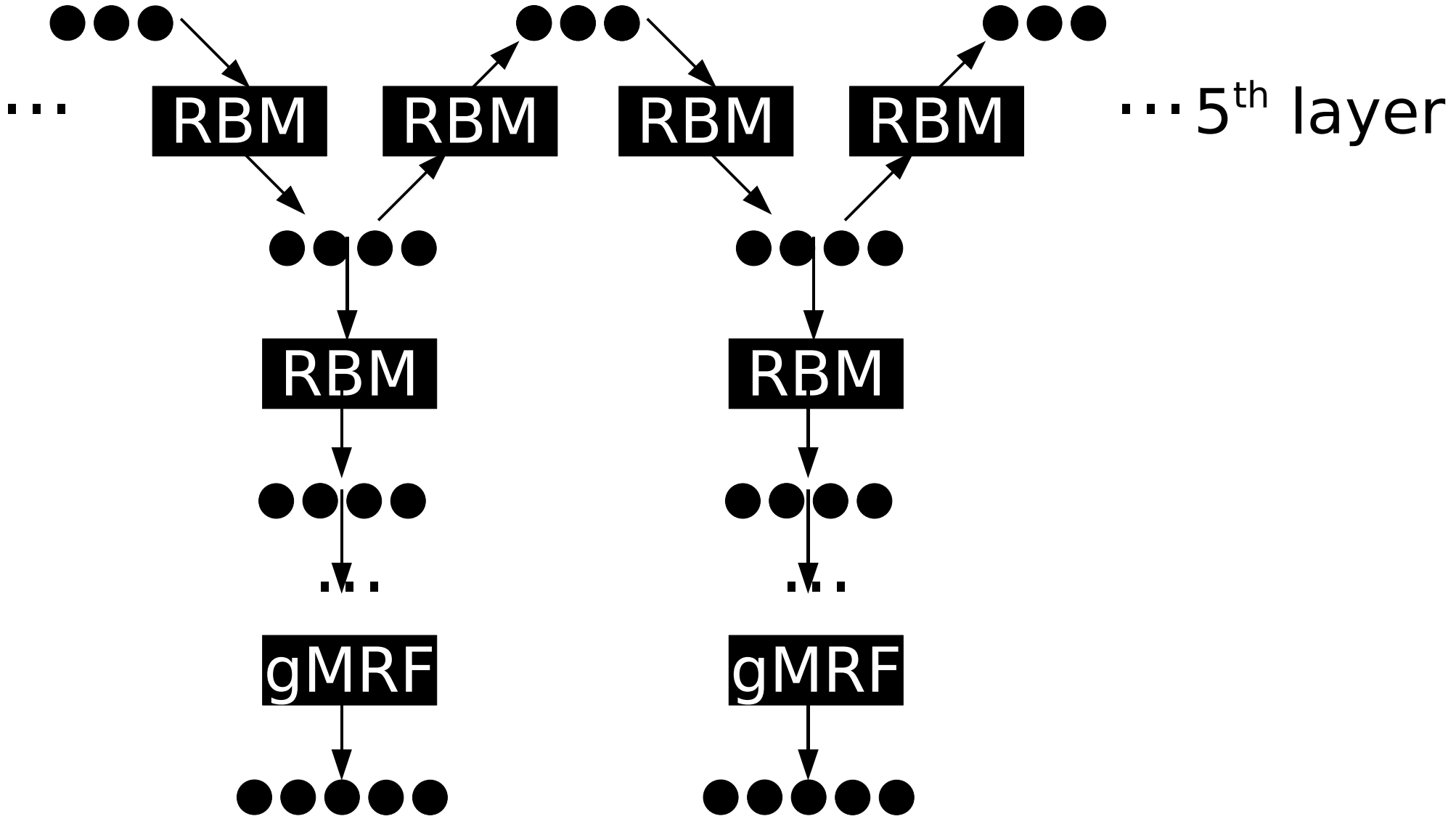
Facial Expression Recognition

- We can draw samples from the model (5th layer with 128 hiddens)



Facial Expression Recognition

- We can draw samples from the model (5th layer with 128 hiddens)



Facial Expression Recognition

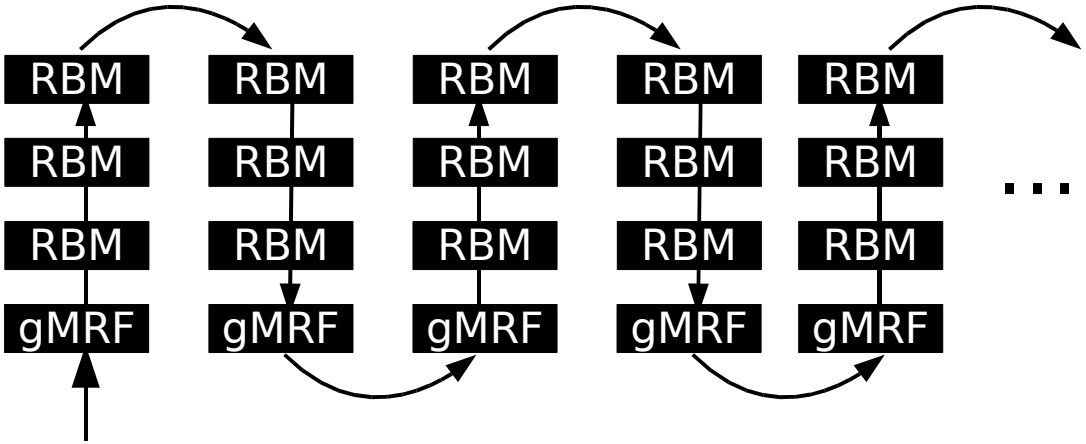
- We can draw samples from the model (5th layer with 128 hiddens)

Facial Expression Recognition

- We can draw samples from the model (5th layer with 128 hiddens)

Facial Expression Recognition

- 7 synthetic occlusions
- use generative model to fill-in
  (conditional on the known pixels)



Facial Expression Recognition

originals

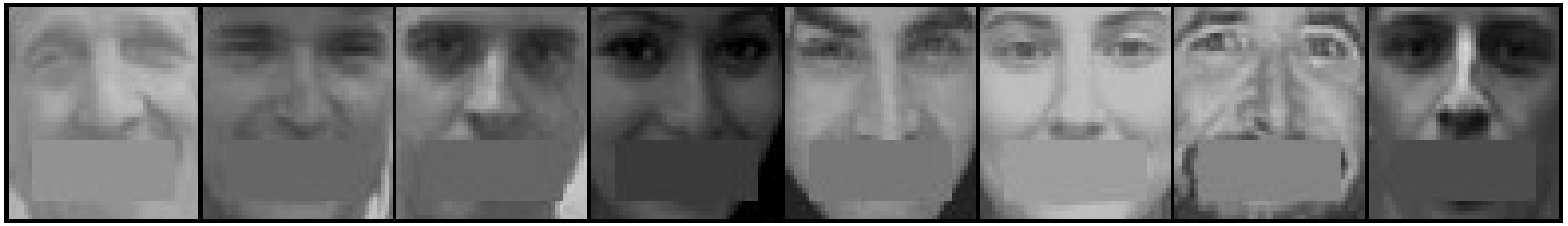
Type 1 occlusion: eyes

Restored images

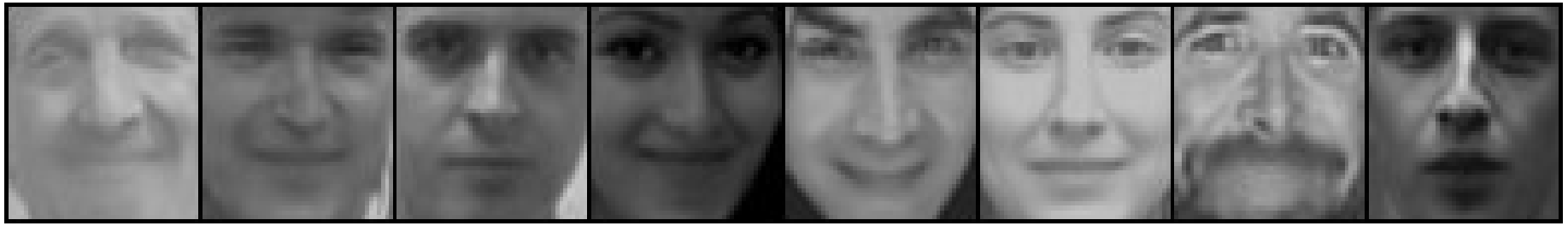
Facial Expression Recognition

originals

Type 2 occlusion: mouth



Restored images



Facial Expression Recognition

originals

Type 3 occlusion: right half

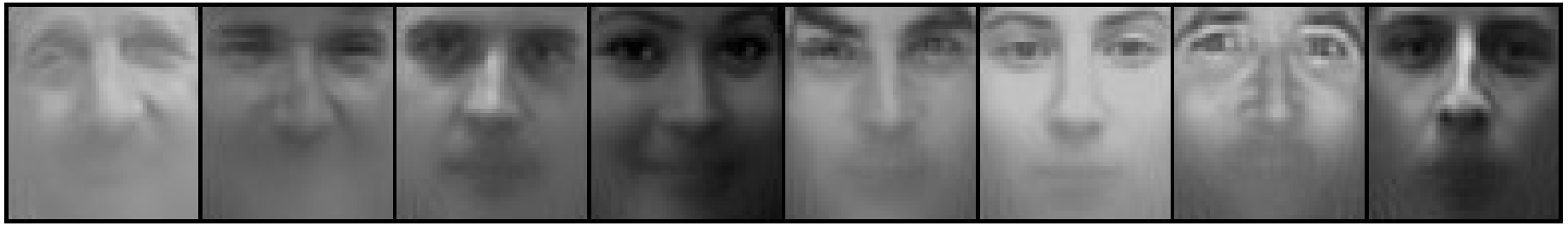
Restored images

Facial Expression Recognition

originals

Type 4 occlusion: bottom half

Restored images

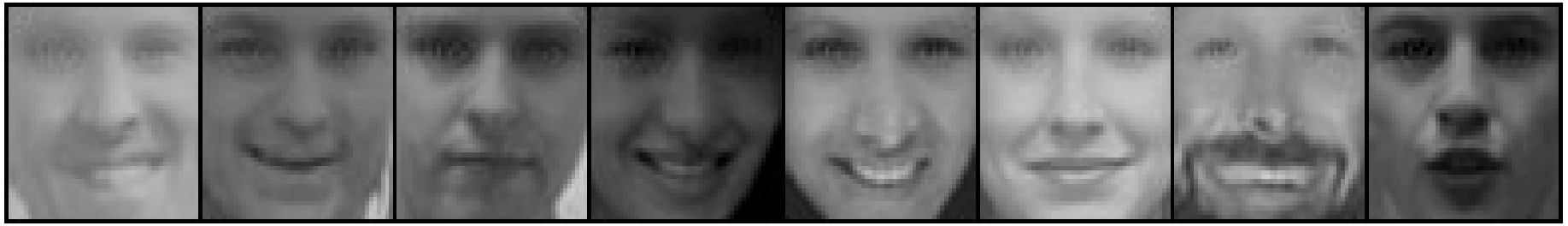


Facial Expression Recognition

originals

Type 5 occlusion: top half

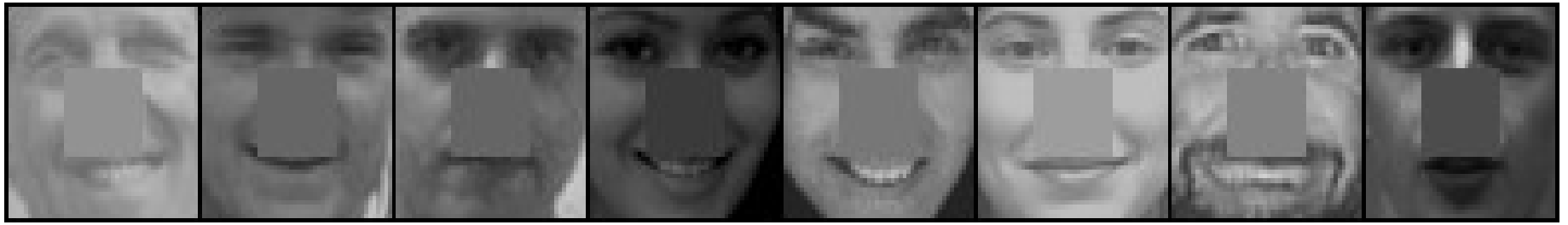
Restored images



Facial Expression Recognition

originals

Type 6 occlusion: nose

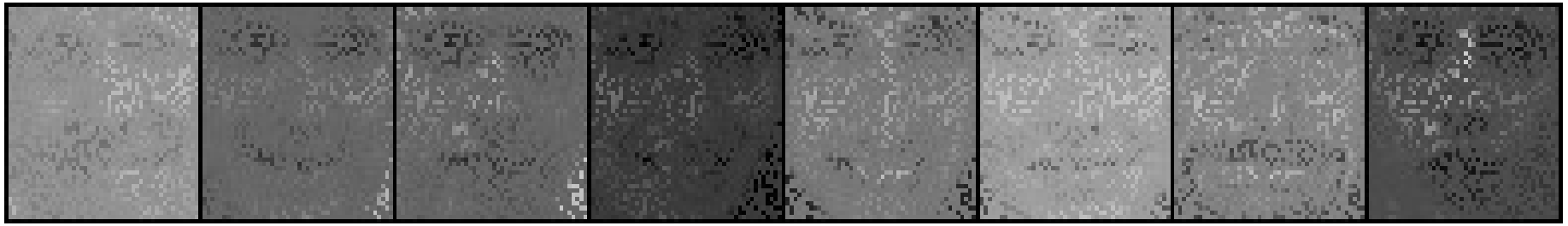


Restored images

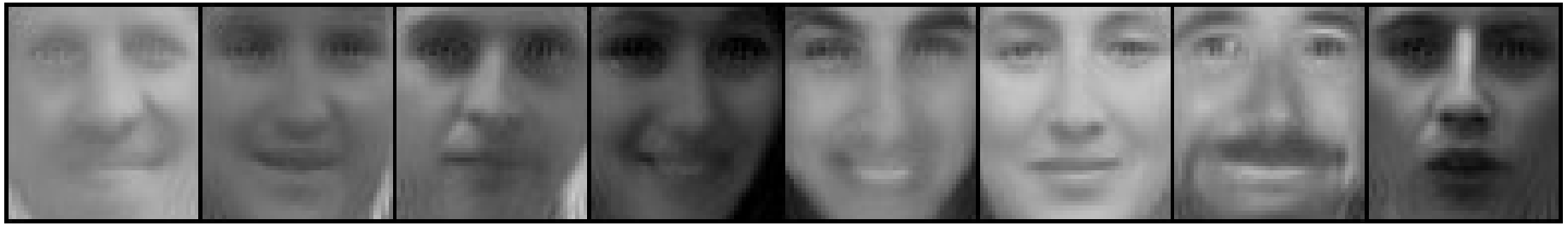
Facial Expression Recognition

originals

Type 7 occlusion: 70% of pixels at random



Restored images



Facial Expression Recognition

Origina

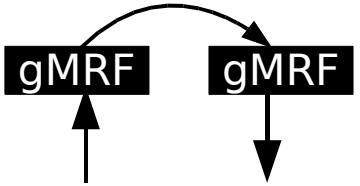
Input

Facial Expression Recognition

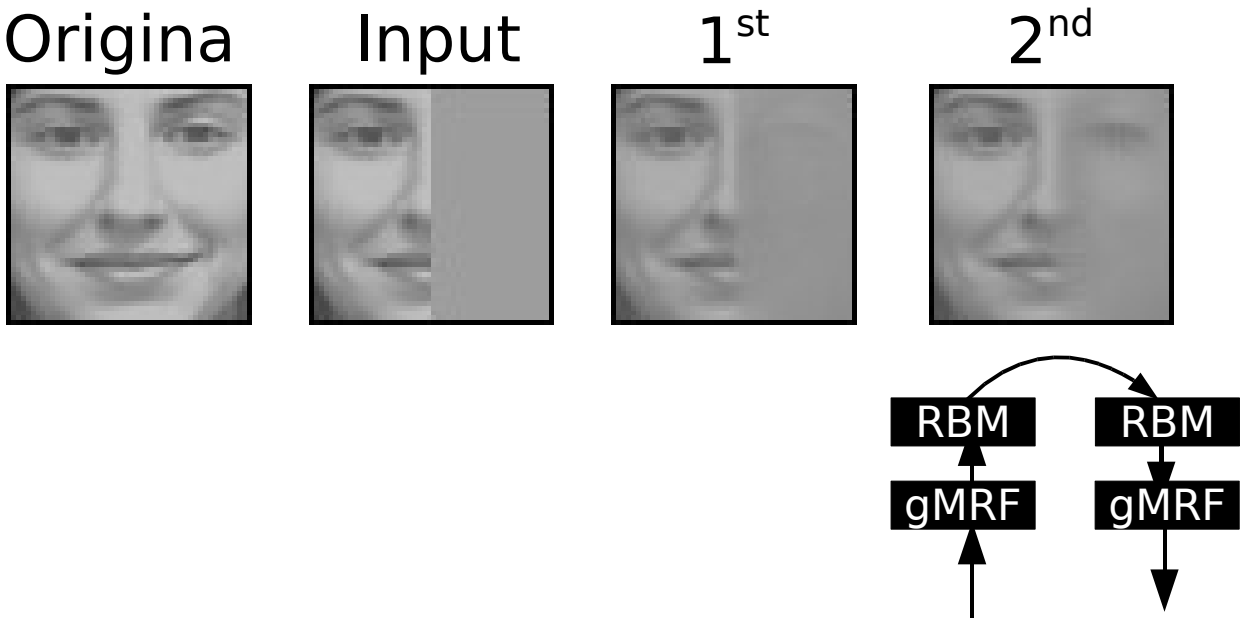
Origina

Input

1st

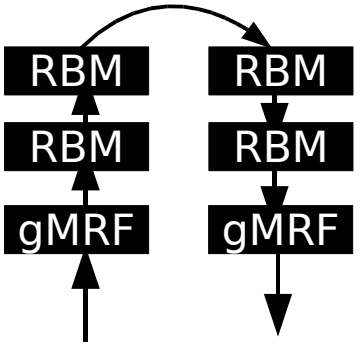


Facial Expression Recognition

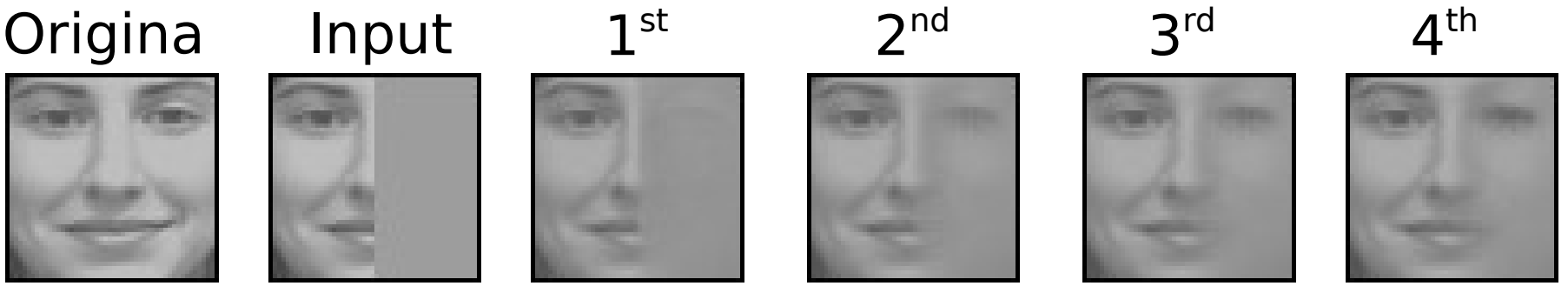


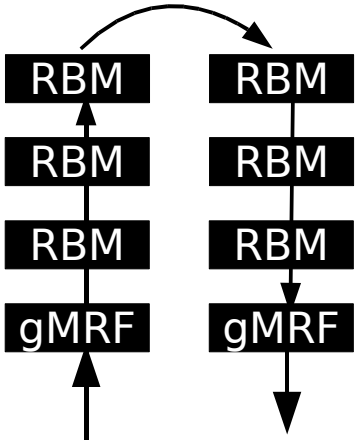
Facial Expression Recognition

Origina Input 1st 2nd 3rd

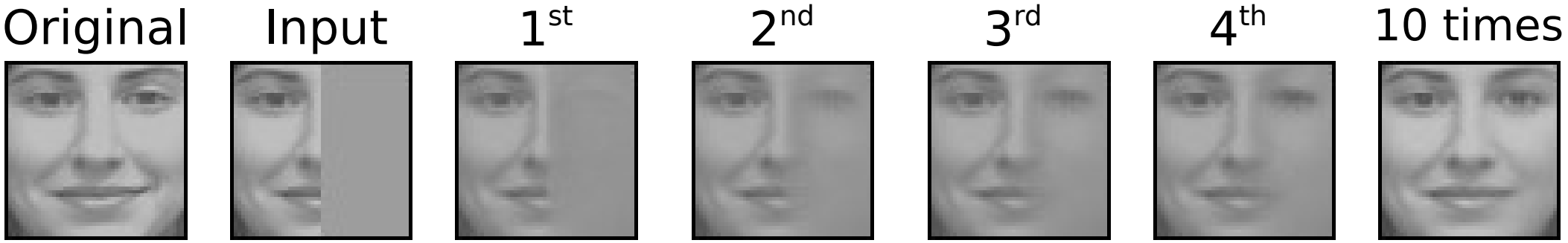


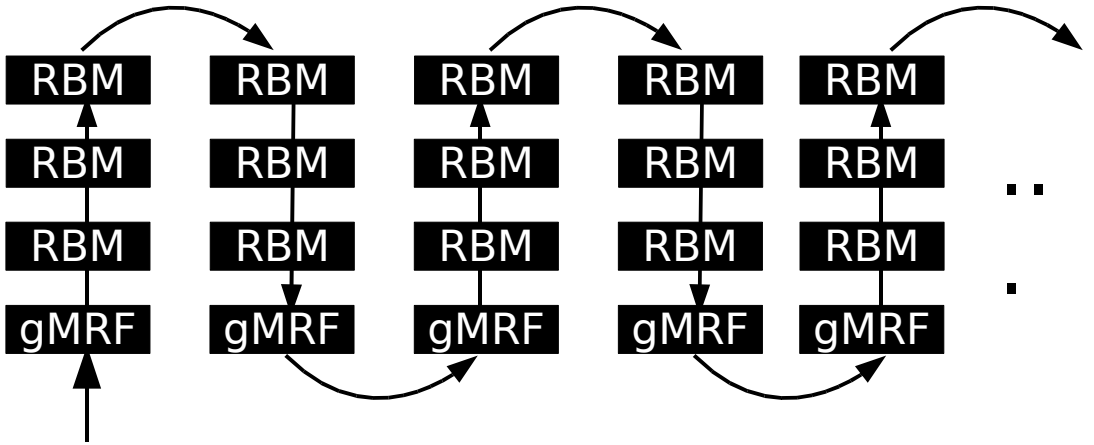
Facial Expression Recognition





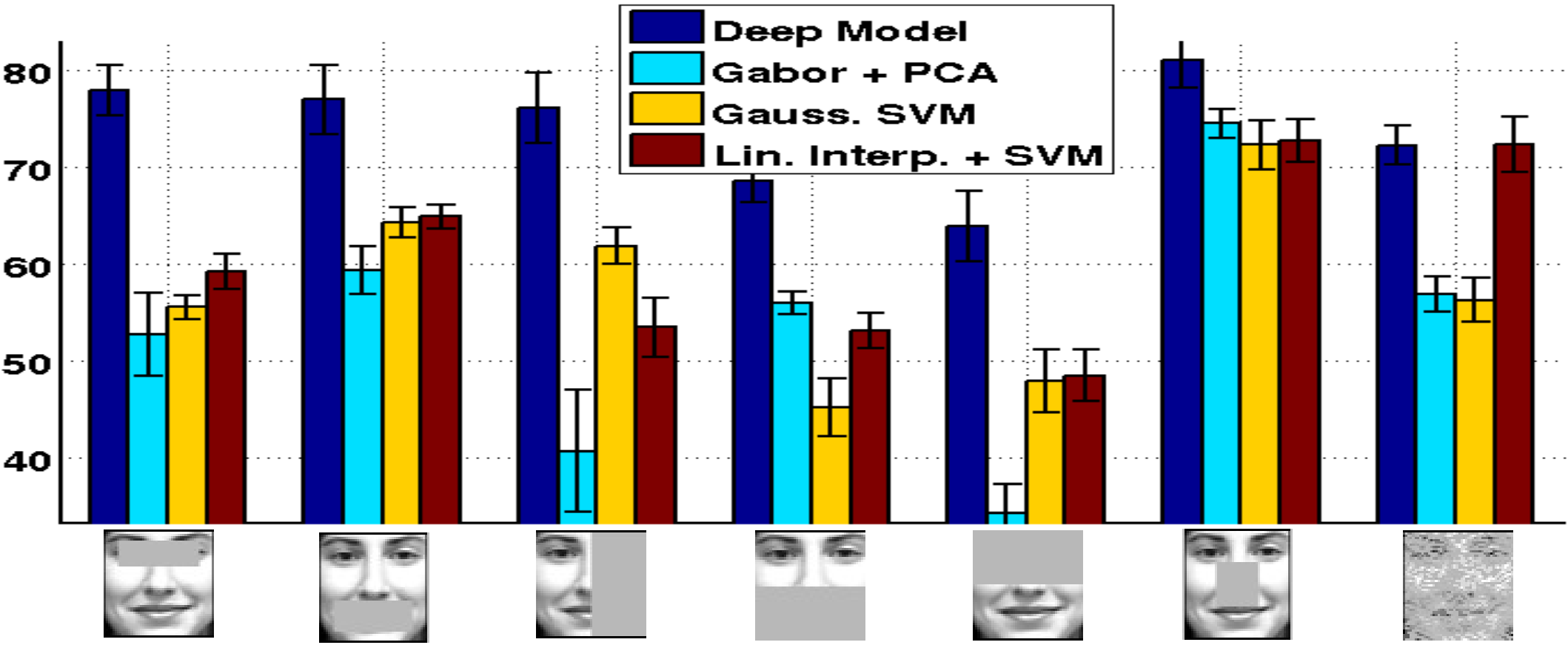
Facial Expression Recognition





Facial Expression Recognition

CASE 1: original images for training, occluded for test



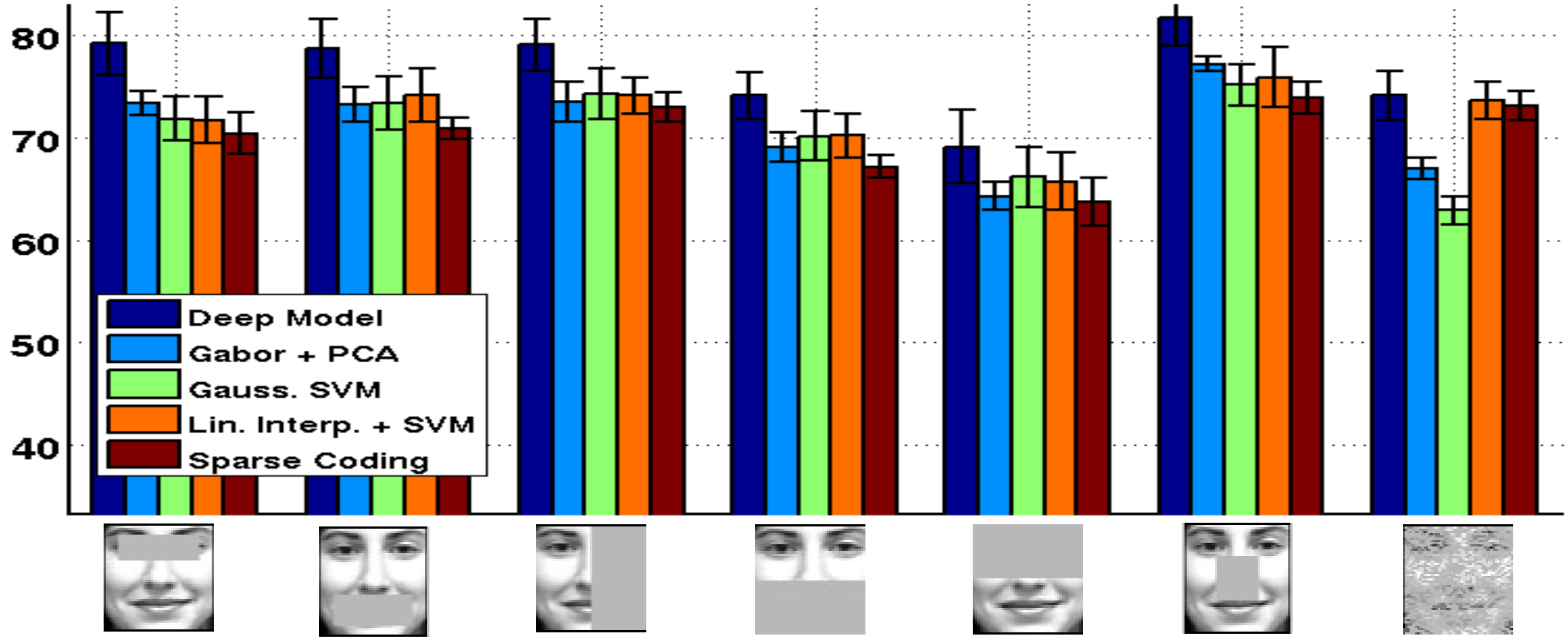
Dailey, et al. J. Cog. Neuros. 2003

Wright, et al. PAMI 2008

Ranzato, et al. CVPR 2011

Facial Expression Recognition

CASE 2: occluded images for both training and test



Dailey, et al. J. Cog. Neuros. 2003

Wright, et al. PAMI 2008

Ranzato, et al. CVPR 2011

Outline

- mathematical formulation of the model
- training
- learning acoustic features for spech recognition
- generation of natural images
- recognition of facial expression under occlusion
- conclusion

Summary

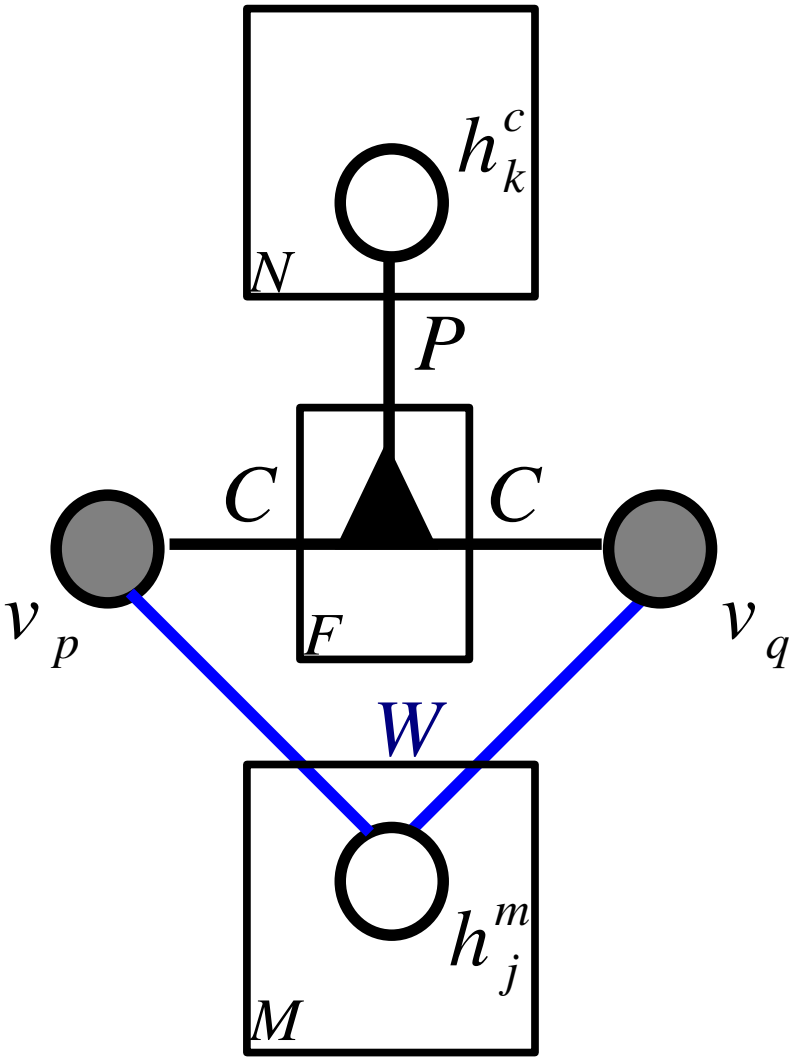
- Unsupervised Learning: regularize & learn representations

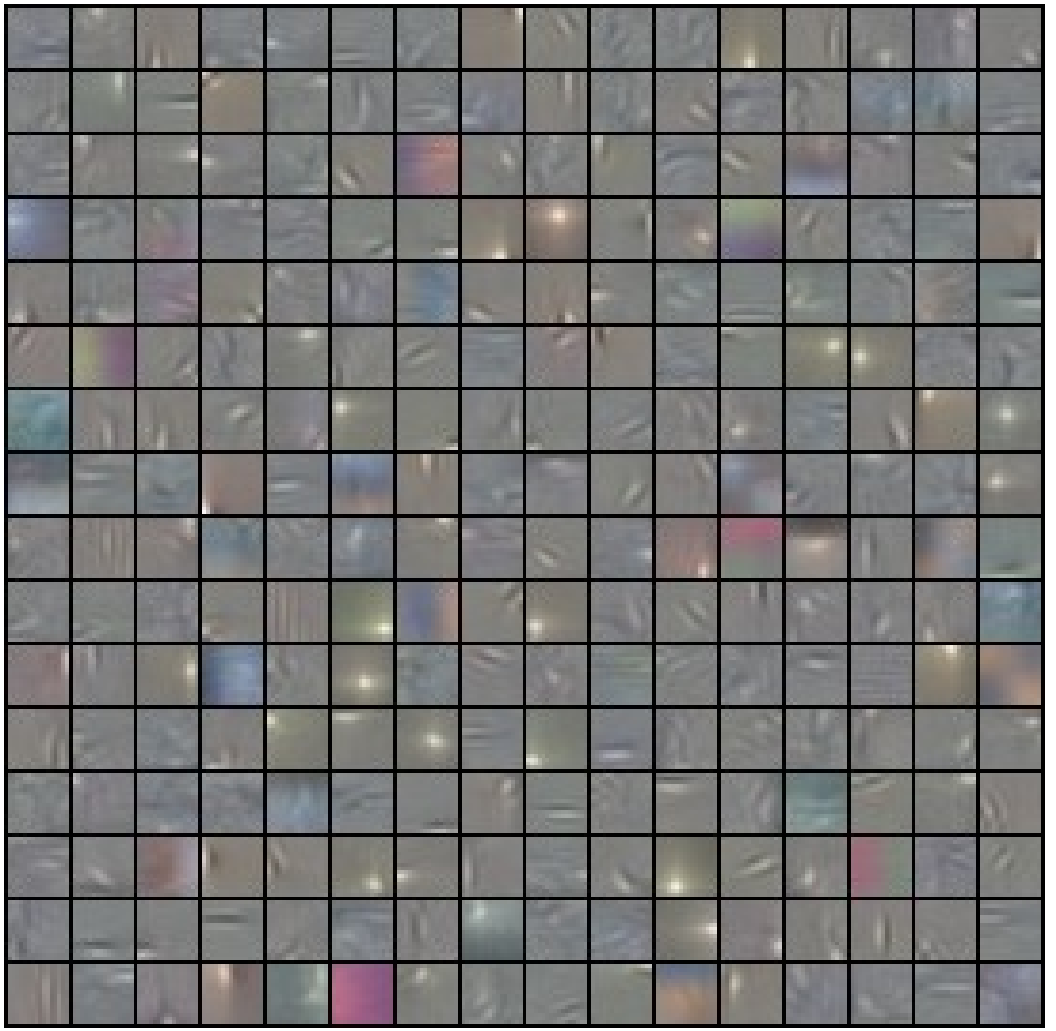
- Deep Generative Model
  - 1st layer: gated MRF
  - Higher layers: binary RBM's
  - fast inference

- Realistic generation: natural images
- Applications:
  - facial expression recognition robust to occlusion
  - speech recognition

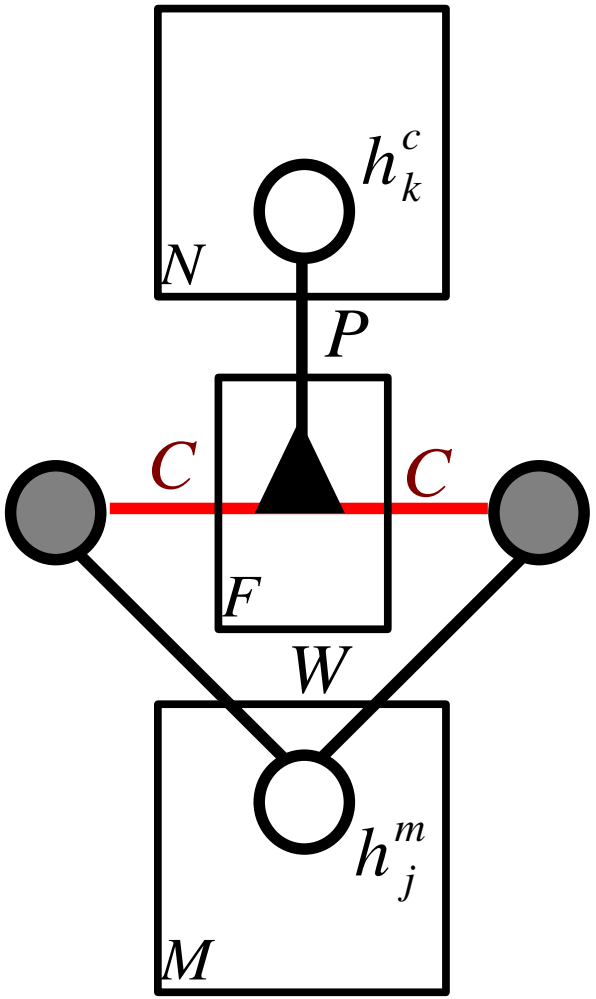
THANK YOU

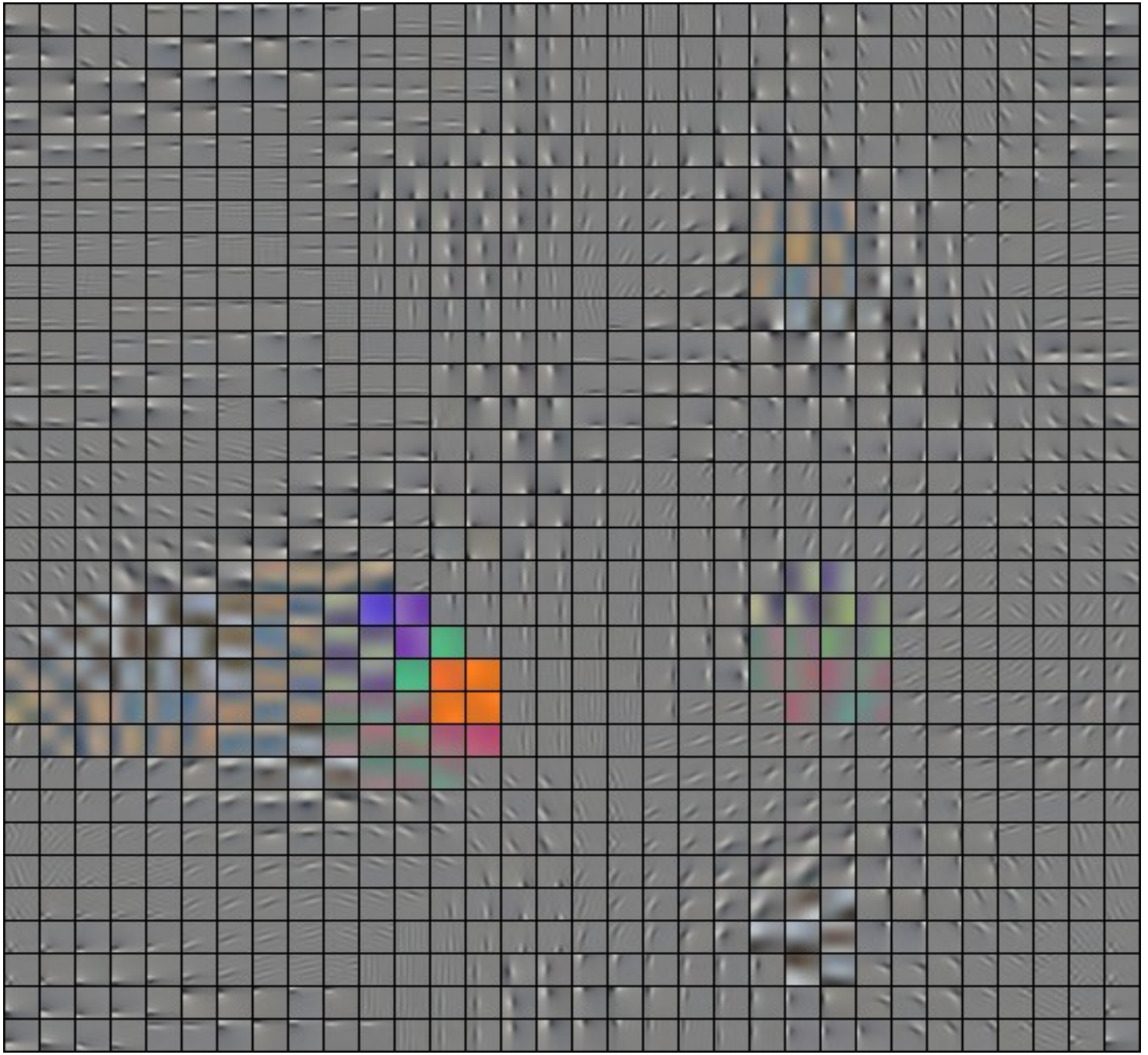
Learned Filters





Learned Filters



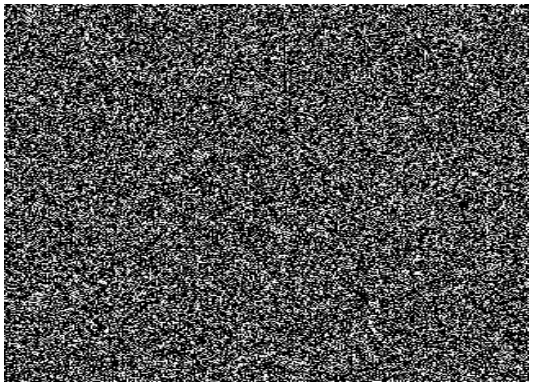


Using -Energy to Score Images

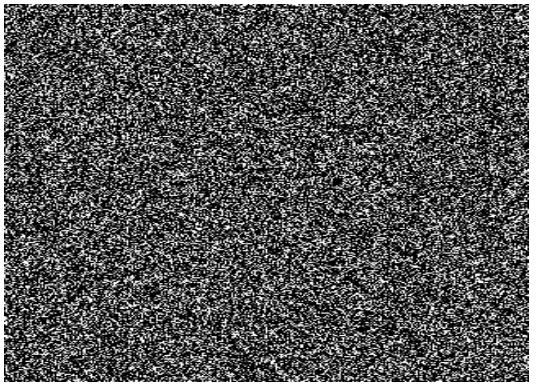
less likely

test images

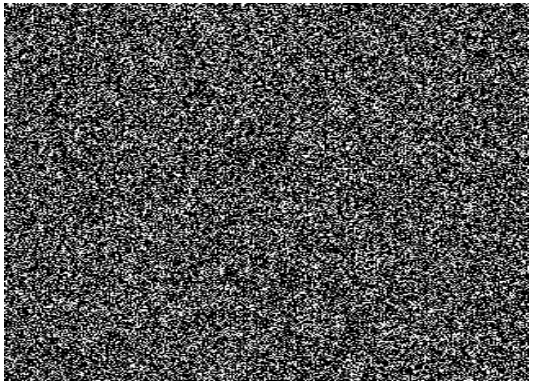
>



>



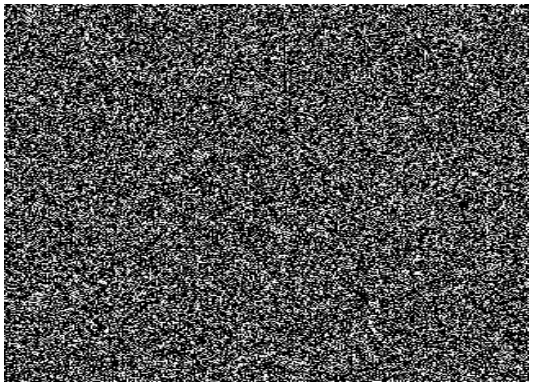
>

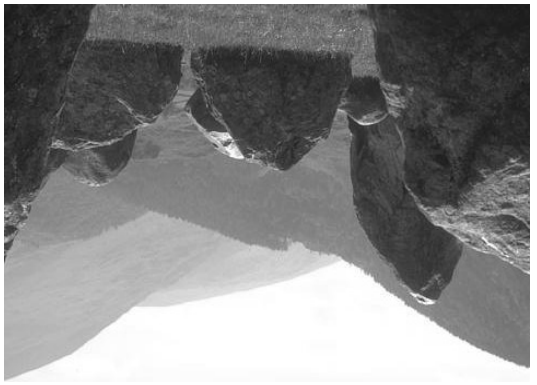
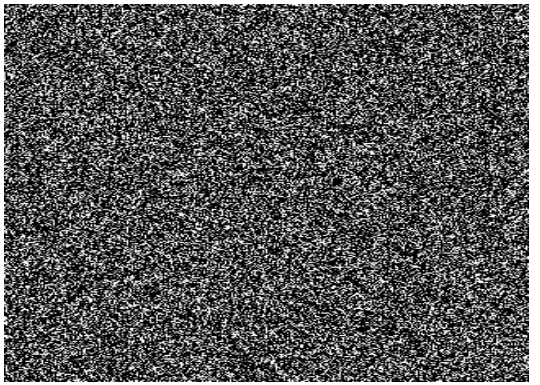


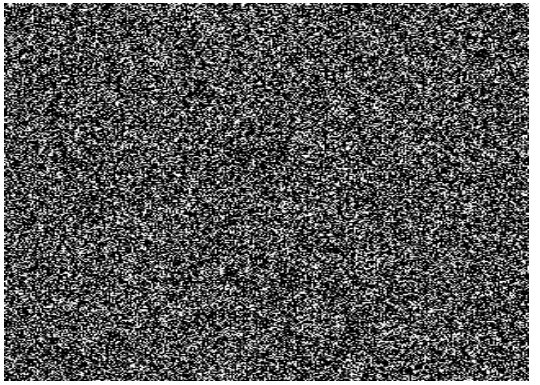
Using Energy to Score Images

Upside-down images





Using Energy to Score Images

Average of those images for which difference of energy is higher