Deep Learning Methods for Vision

Rob Fergus

Dept. of Computer Science,
Courant Institute,
New York University

Overview

- Learning Feature Hierarchies for Vision
  – Mainly for recognition

- Many possible titles:
  – Deep Learning
  – Feature Learning

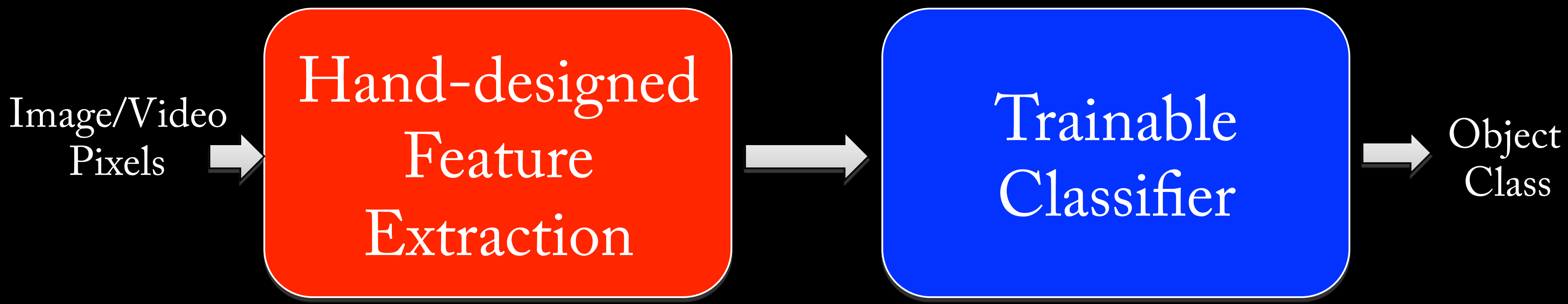
- This talk: Basic concepts
  Links to existing vision approaches

Overview

- Learning Feature Hierarchies for Vision
  - For object recognition

- This talk: Basic concepts
  Links to existing vision approaches

Existing Recognition Approach



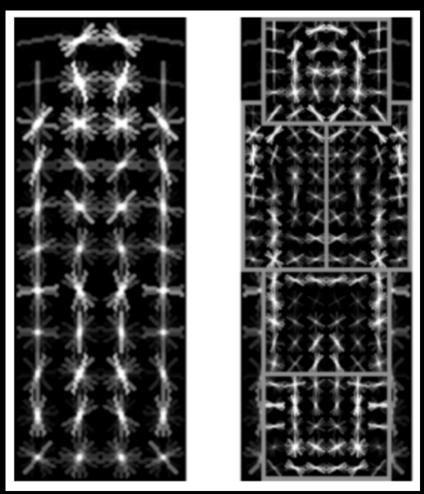
- Features are not learned

- Trainable classifier is often generic (e.g. SVM)

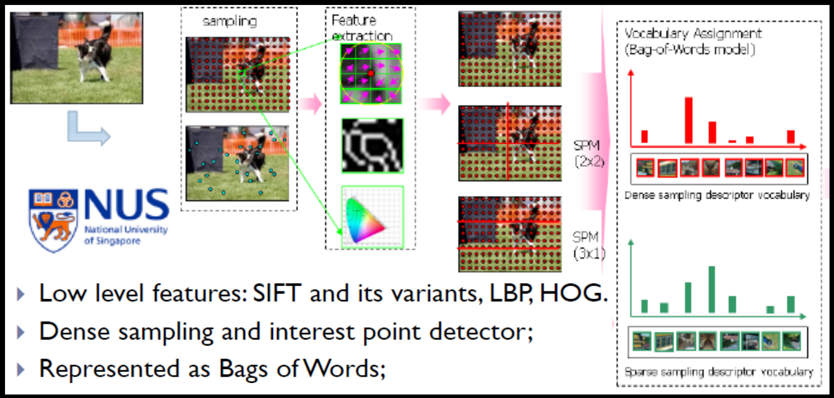
Slide: Y.LeCun

Motivation

- Features are key to recent progress in recognition
- Multitude of hand-designed features currently in use
  - SIFT, HOG, LBP, MSER, Color-SIFT…………
- Where next? Better classifiers? Or keep building more features?



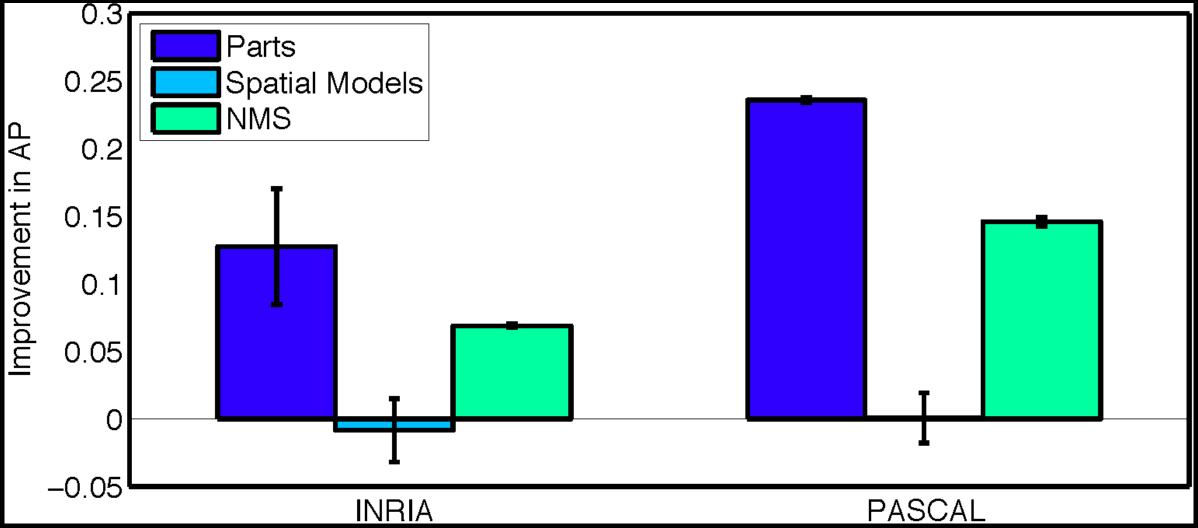
Felzenszwalb, Girshick,
McAllester and Ramanan, PAMI 2007



Yan & Huang
(Winner of PASCAL 2010 classification competition)

What Limits Current Performance?

- Ablation studies on Deformable Parts Model
  - Felzenszwalb, Girshick, McAllester, Ramanan, PAMI'10
- Replace each part with humans (Amazon Turk):

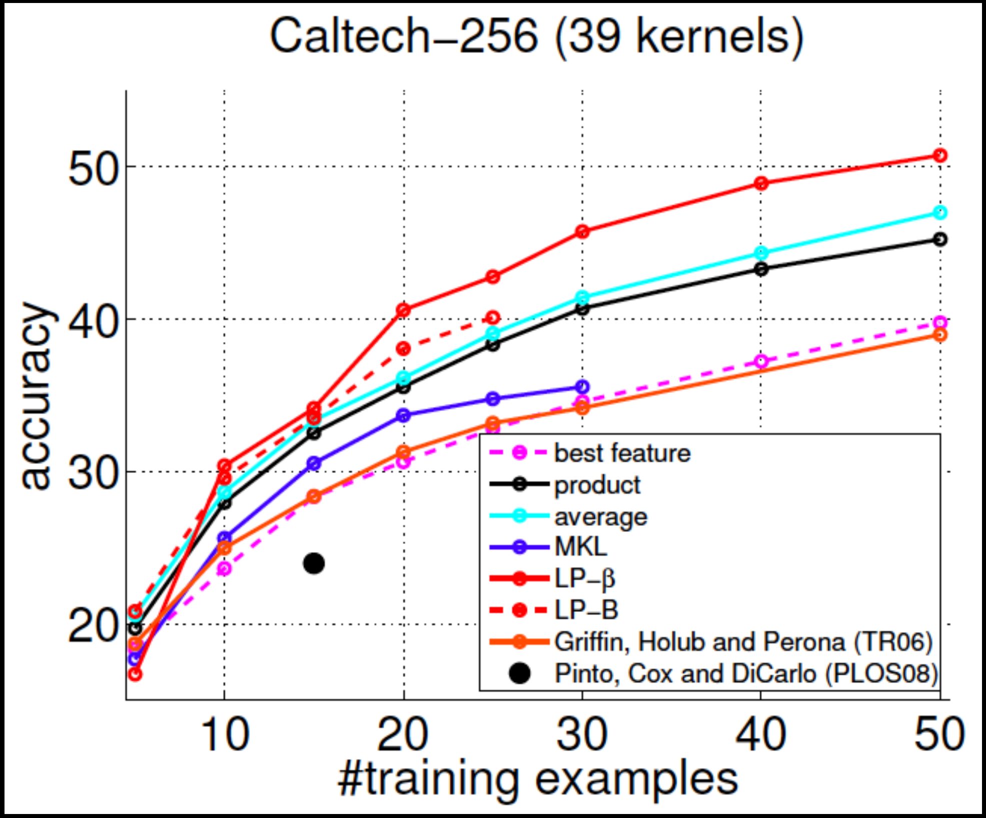


Parikh & Zitnick, CVPR'10

Hand-Crafted Features

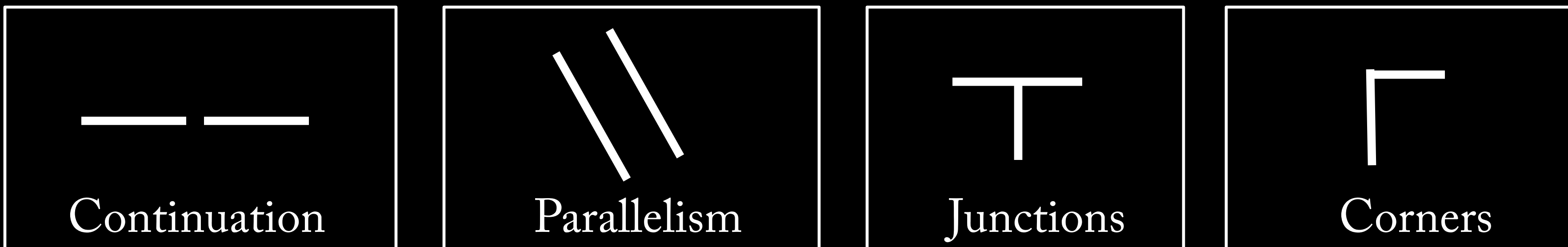
- LP-β Multiple Kernel Learning
  - Gehler and Nowozin, On Feature Combination for Multiclass Object Classification, ICCV'09
- 39 different kernels
  - PHOG, SIFT, V1S+, Region Cov. Etc.
- MKL only gets few % gain over averaging features

→ Features are doing the work



Mid-Level Representations

- Mid-level cues



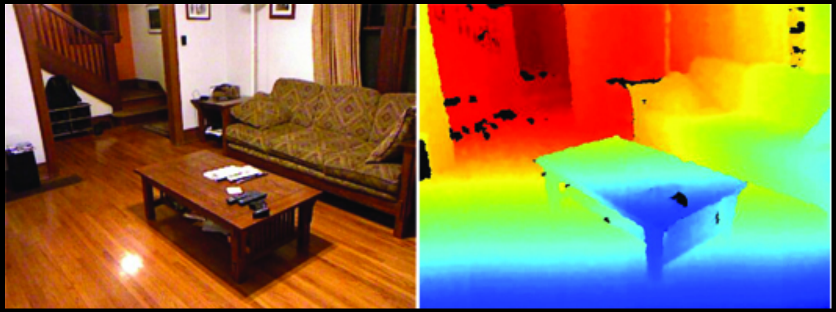
"Tokens" from Vision by D.Marr: 

- Object parts:

- Difficult to hand-engineer → What about learning them?

Why Learn Features?

- Better performance
- Other domains (unclear how to hand engineer):
  - Kinect
  - Video
  - Multi spectral



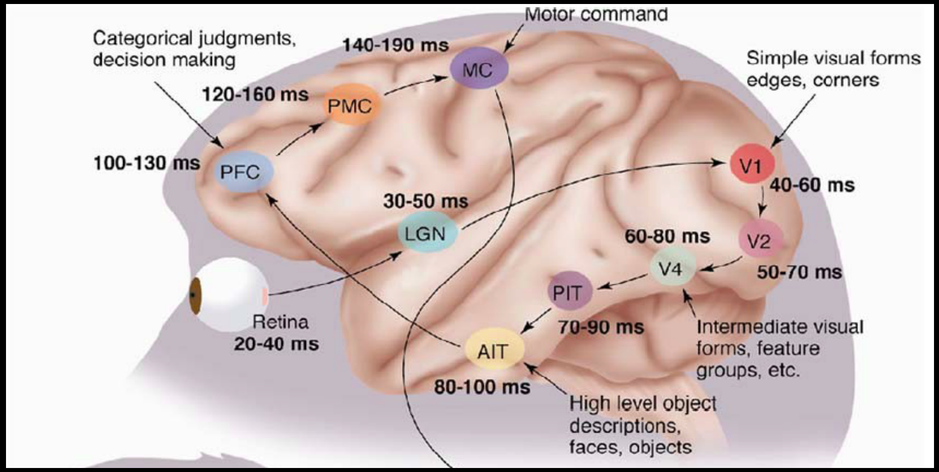
- Feature computation time
  - Dozens of features now regularly used
  - Getting prohibitive for large datasets (10's sec /image)

Why Hierarchy?

Theoretical:

"…well-known depth-breadth tradeoff in circuits design [Hastad 1987]. This suggests many functions can be much more efficiently represented with deeper architectures…" [Bengio & LeCun 2007]

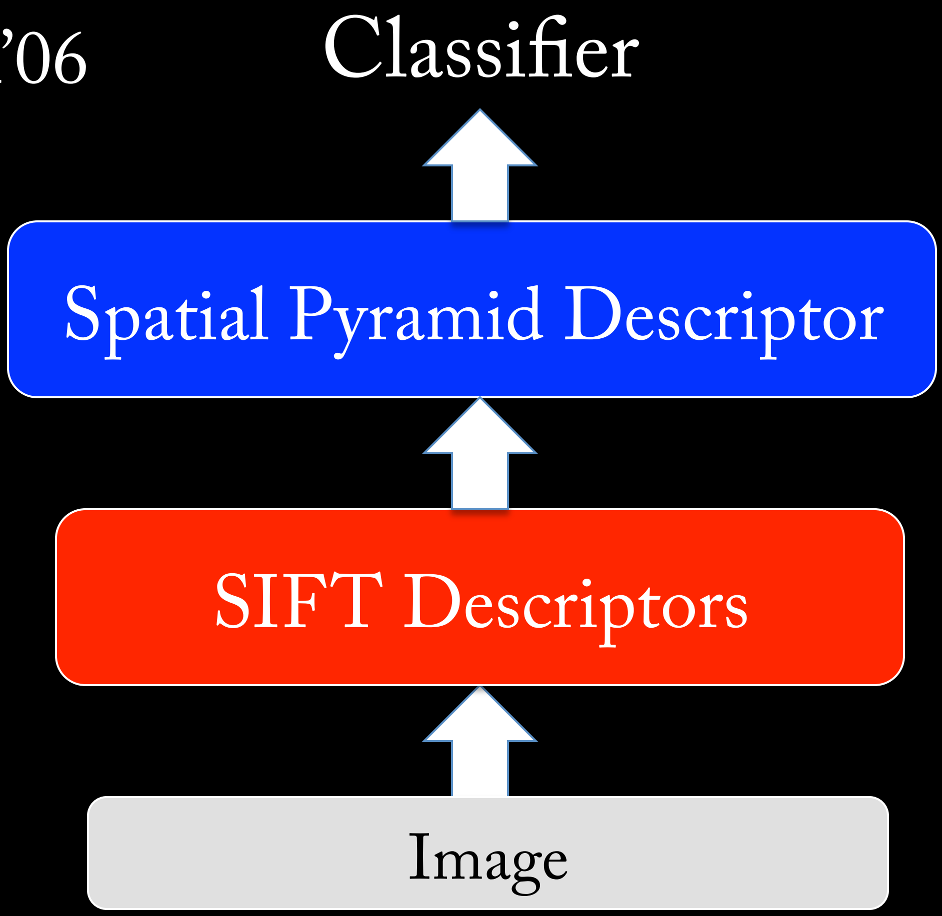
Biological: Visual cortex is hierarchical



[Thorpe]

Hierarchies in Vision

- Spatial Pyramid Matching
  - Lazebnik et al. CVPR'06
- 2 layer hierarchy
  - Spatial Pyramid Descriptor pools VQ'd SIFT



Hierarchies in Vision

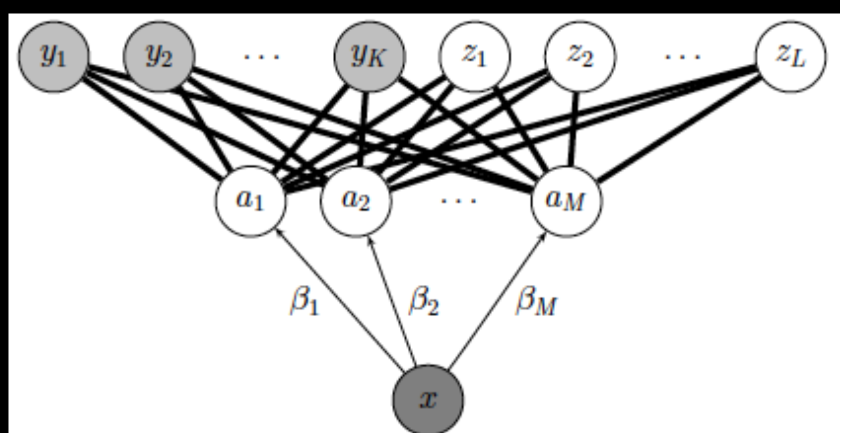
- Lampert et al. CVPR'09

- Learn attributes, then classes as combination of attributes

Class Labels

Attributes

Image Features



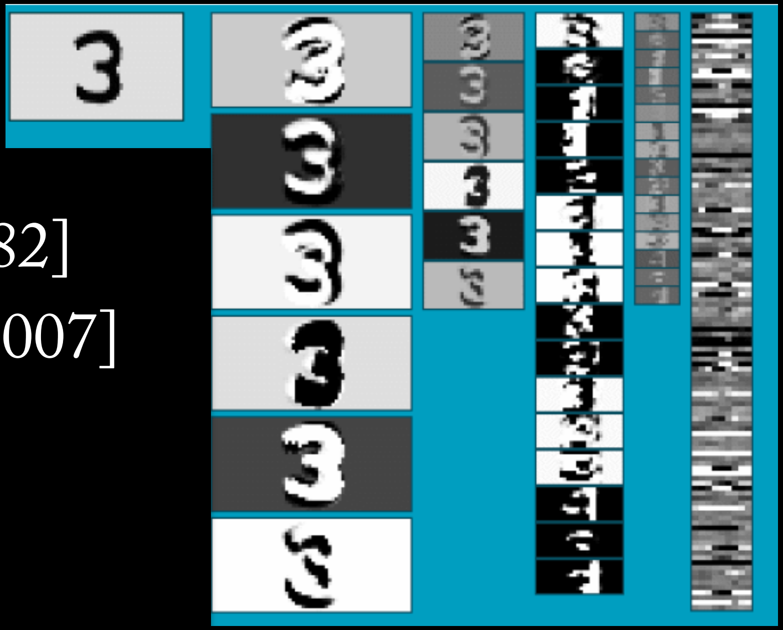
Learning a Hierarchy of Feature Extractors

- Each layer of hierarchy extracts features from output of previous layer
- All the way from pixels → classifier
- Layers have the (nearly) same structure

- Train all layers jointly

Multistage Hubel-Wiesel Architecture

- Stack multiple stages of simple cells / complex cells layers
- Higher stages compute more global, more invariant features
- Classification layer on top



History:

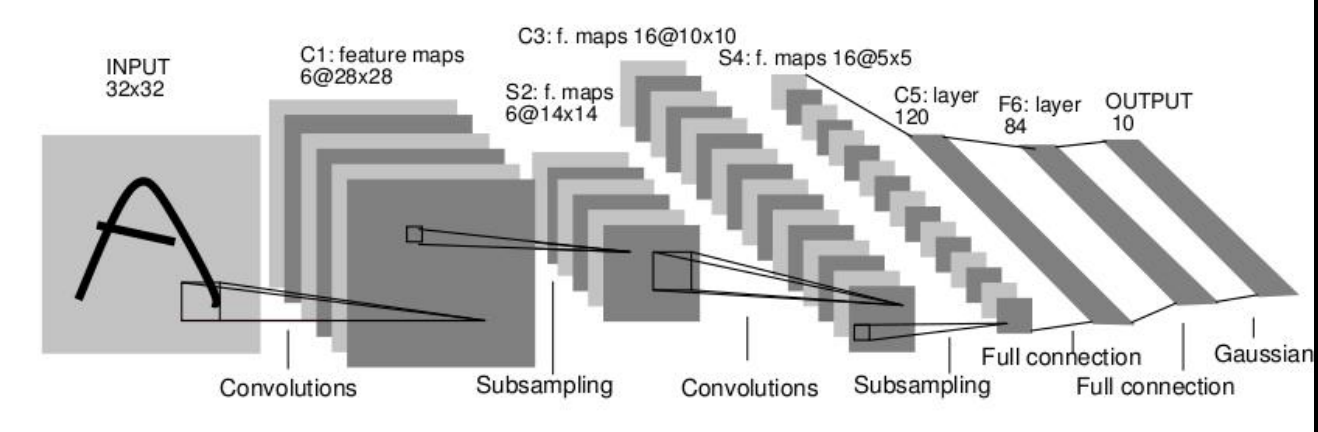
- Neocognitron [Fukushima 1971-1982]
- Convolutional Nets [LeCun 1988-2007]
- HMAX [Poggio 2002-2006]
- Many others…

QUESTION: How do we find (or learn) the filters?

Slide: Y.LeCun

Supervised Learning

- Convolutional Neural Networks
  - Back-propagation



[LeCun *et al.* 1998]

- Problems:
  - Can be difficult to train deep models:
    - Vanishing gradients
    - Highly non-convex (local minima)
  - Getting enough labels

Unsupervised Learning

- Model distribution of input data
- Can use unlabeled data (unlimited)
- Refine with standard supervised techniques (e.g. backprop)

Single Layer Architecture

Input: Image Pixels / Features

Layer n

Details in the boxes matter

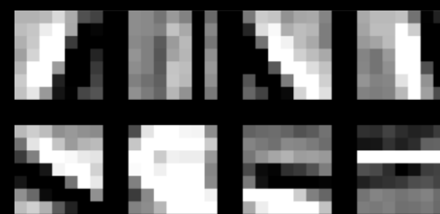
(especially in a hierarchy)

Output: Features / Classifier

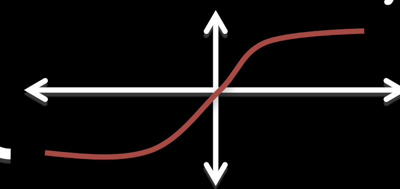
Example Feature Learning Architectures

Pixels / Features

Filter with Dictionary (patch/tiled/convolutional)



+ Non-linearity



Normalization between feature responses

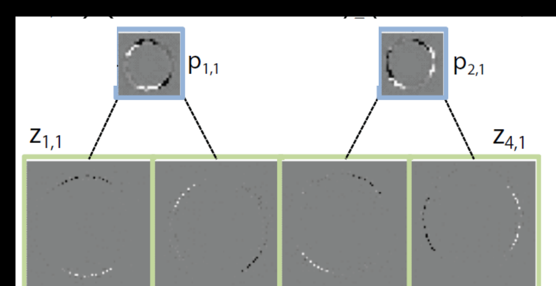
(Group) ℓ_1 Sparsity

Max / Softmax

Local Contrast Normalization (Subtractive & Divisive)

Not an exact separation

Spatial/Feature (Sum or Max)

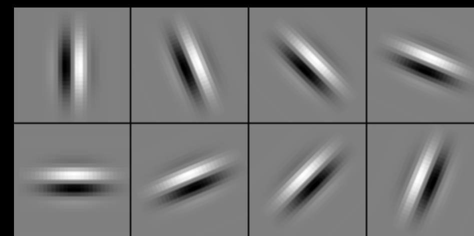


Features

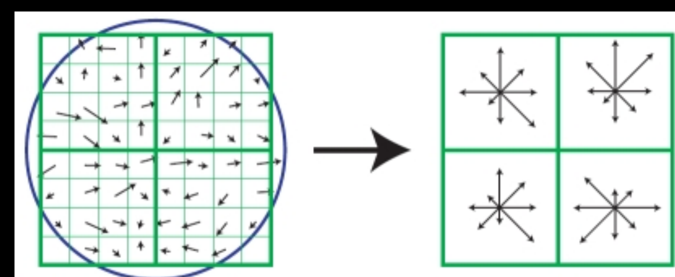
SIFT Descriptor

Image Pixels

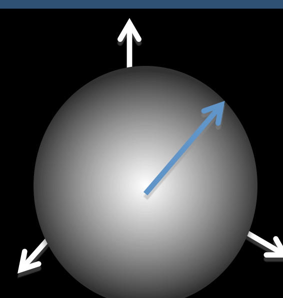
Apply Gabor filters



Spatial pool (Sum)



Normalize to unit length



Feature Vector

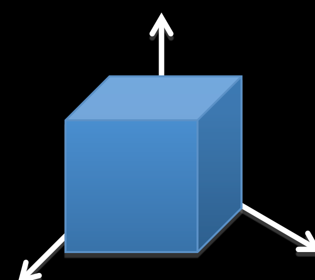
Spatial Pyramid Matching

Lazebnik, Schmid, Ponce [CVPR 2006]

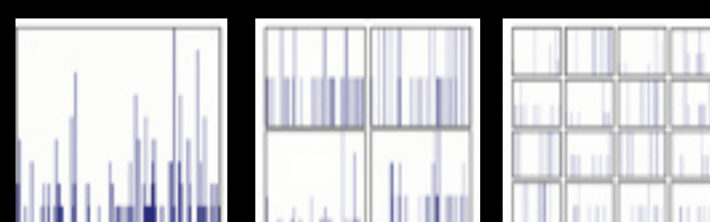
SIFT Features

Filter with Visual Words

Max



Multi-scale spatial pool (Sum)

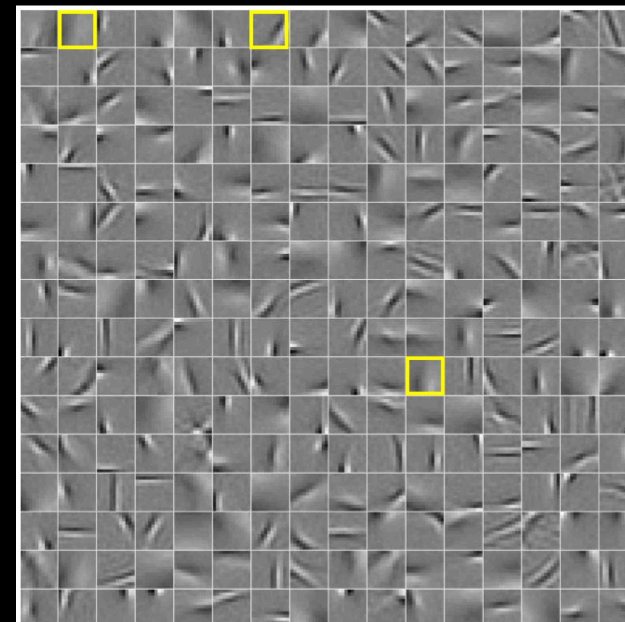


Classifier

Filtering

- Patch
  - Image as a set of patches

Input



Filters

#patches

#filters

Filtering

- Convolutional
  - Translation equivariance
  - Tied filter weights

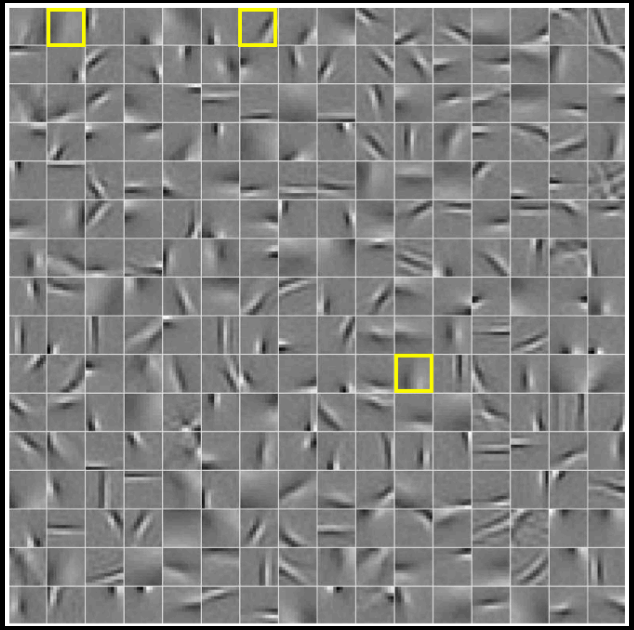
(same at each position → few parameters)

Input

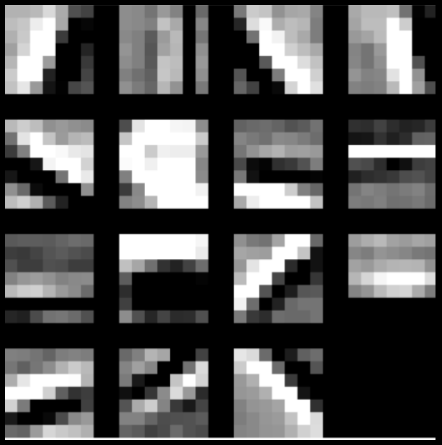
Feature Map

Translation Equivariance

- Input translation → translation of features
  - Fewer filters needed: no translated replications
  - But still need to cover orientation/frequency



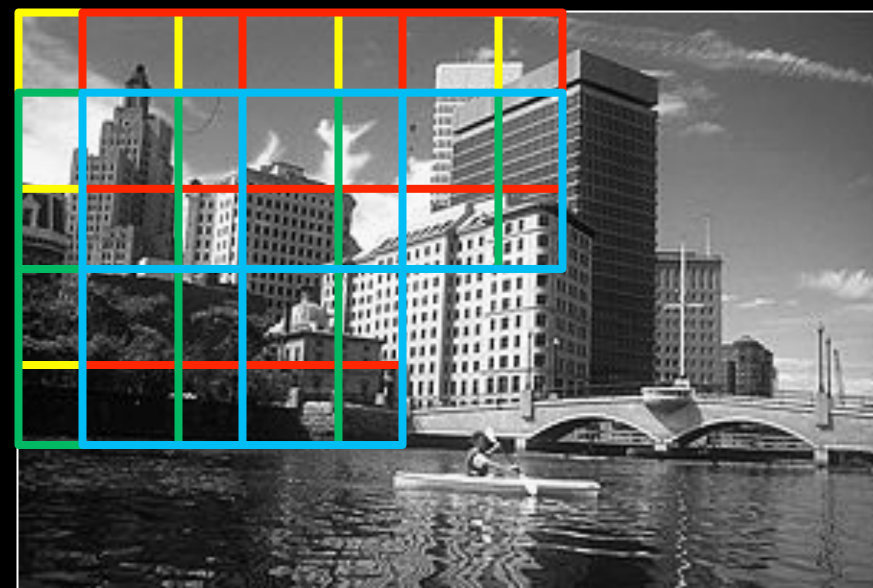
Patch-based



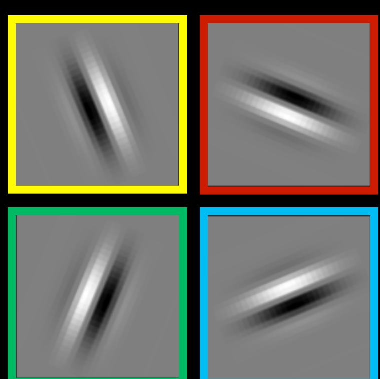
Convolutional

Filtering

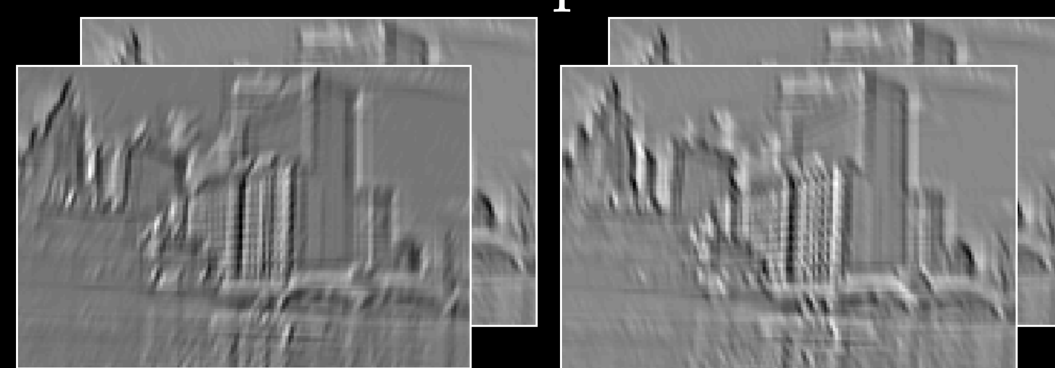
- Tiled
  - Filters repeat every n
  - More filters than convolution for given # features



Input



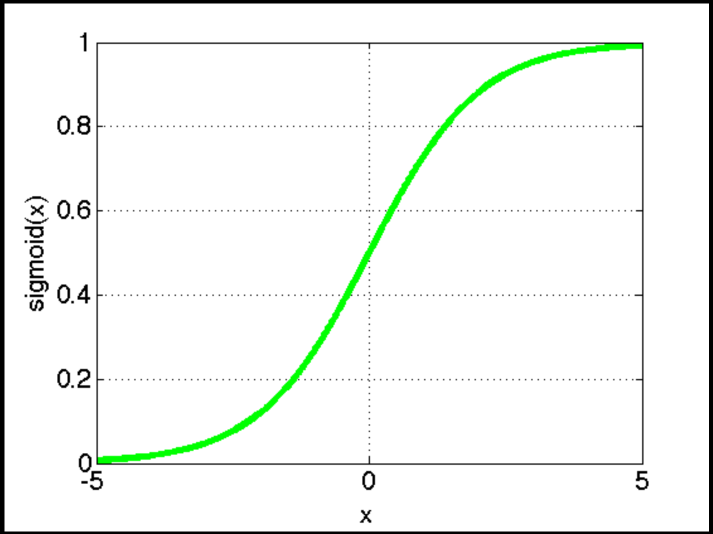
Filters

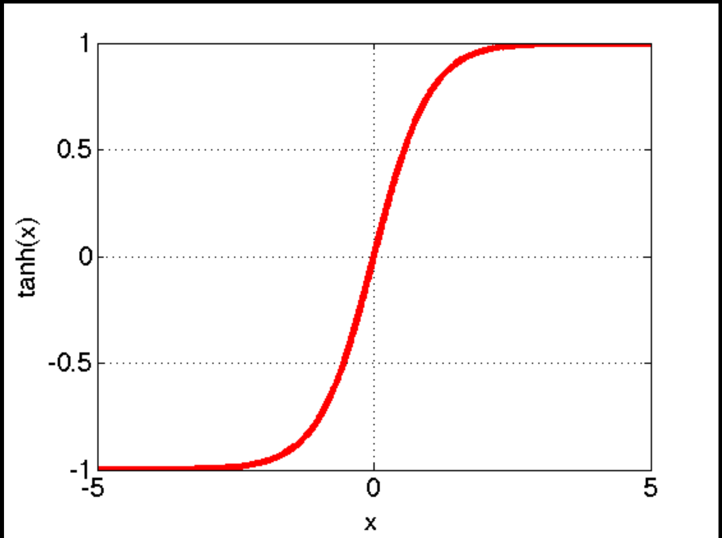


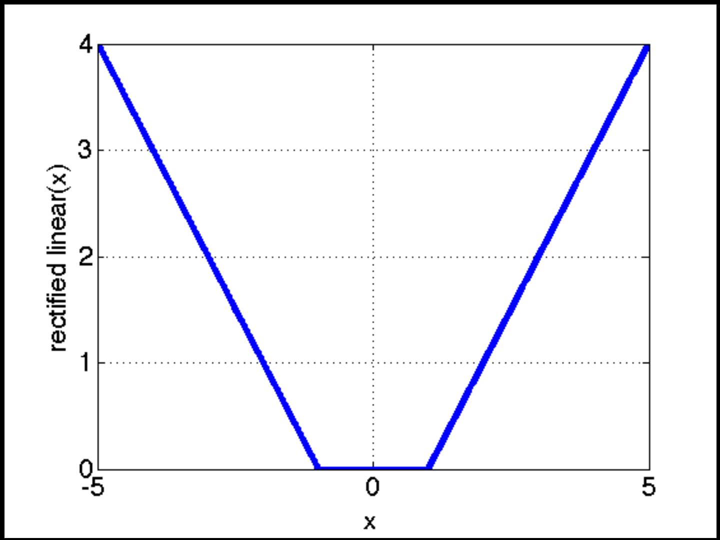
Feature maps

Filtering

- Non-linearity
  - Per-feature independent
  - Tanh
  - Sigmoid: 1/(1+exp(-x))
  - Rectified linear







Normalization

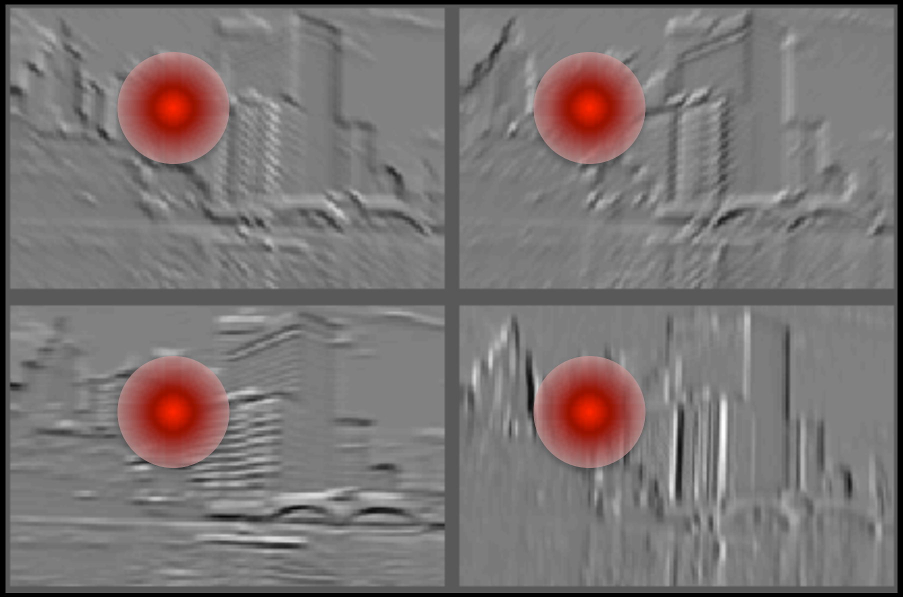
- Contrast normalization
  - See Divisive Normalization in Neuroscience

Input

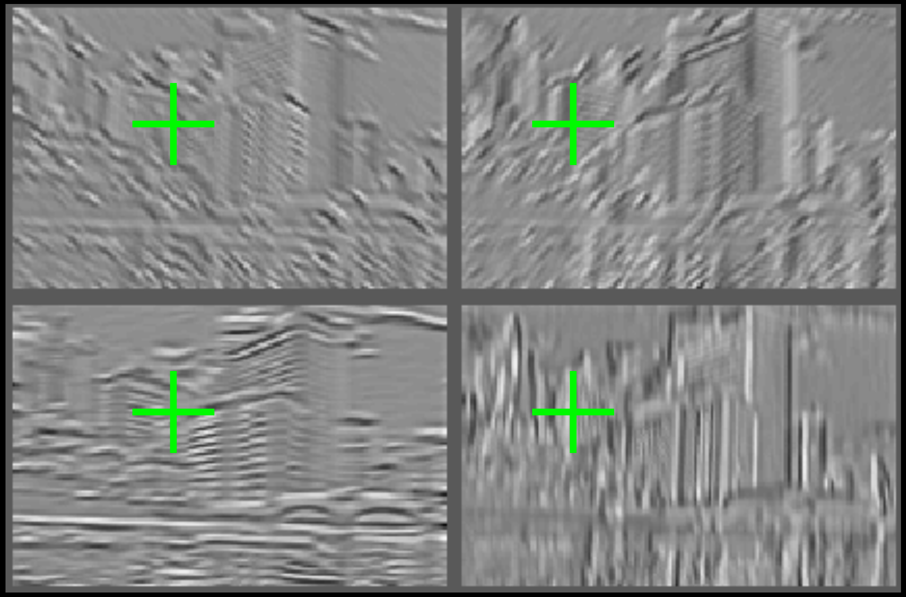
Filters

Normalization

- Contrast normalization (across feature maps)
  - Local mean = 0, local std. = 1, "Local" → 7x7 Gaussian
  - Equalizes the features maps



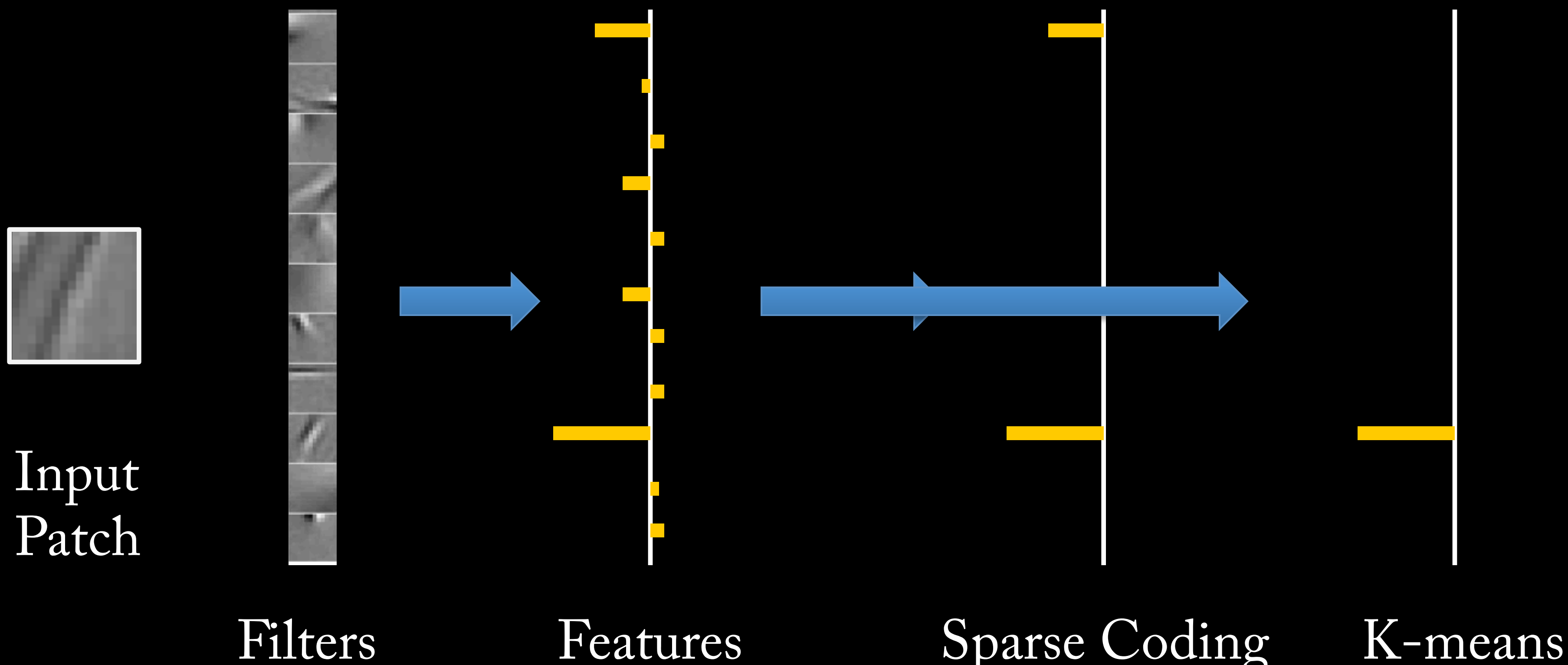
Feature Maps



Feature Maps
After Contrast Normalization

Normalization

- Sparsity
  - Constrain L_0 or L_1 norm of features
  - Iterate with filtering operation (ISTA sparse coding)



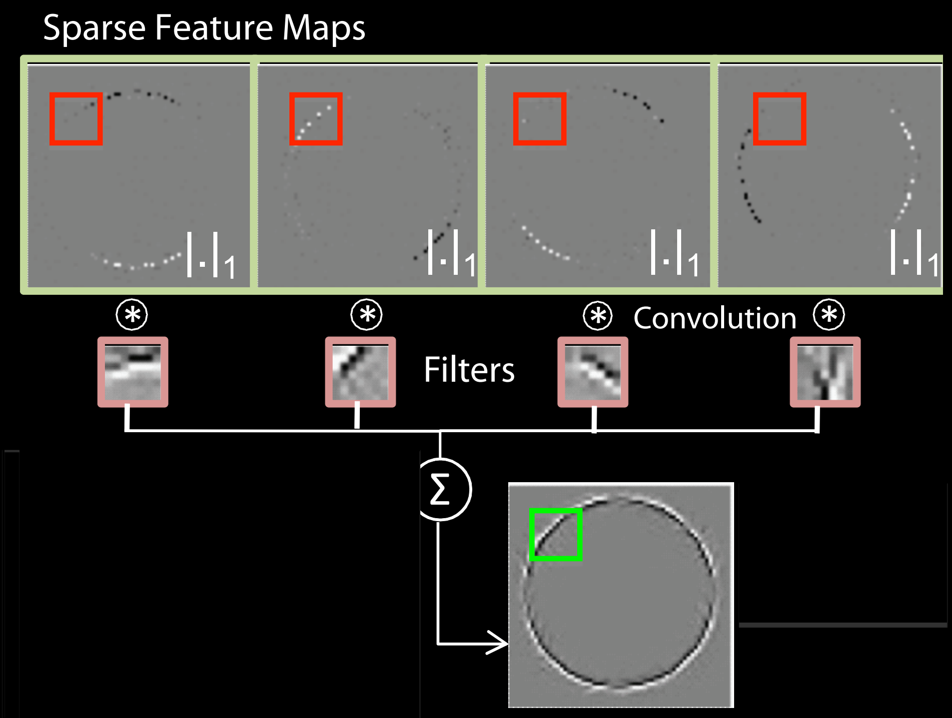
Role of Normalization

- Induces local competition between features to explain input
  - "Explaining away" in graphical models
  - Just like top-down models
  - But more local mechanism

- Filtering alone cannot do this!

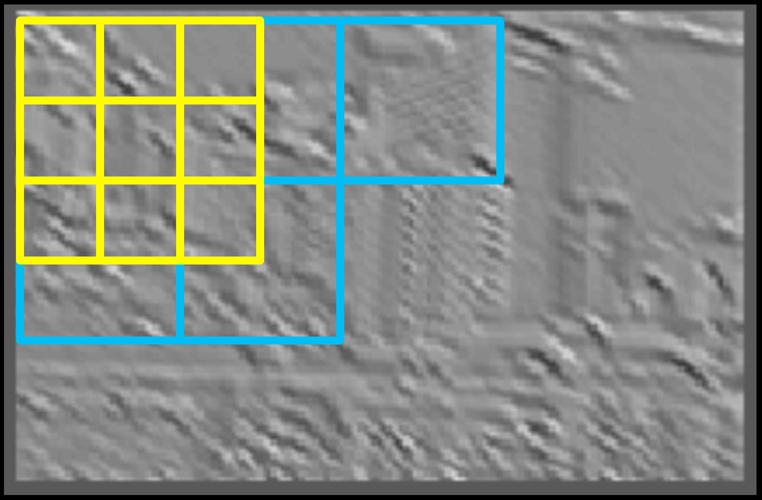
Example:
Convolutional Sparse Coding

from Zeiler et al. [CVPR'10/ICCV'11]

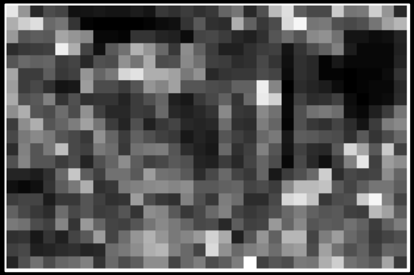


Pooling

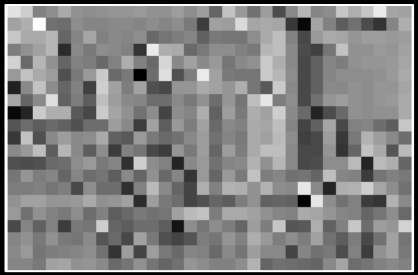
- Spatial Pooling
  - Non-overlapping / overlapping regions
  - Sum or max
  - Boureau et al. ICML'10 for theoretical analysis



Max

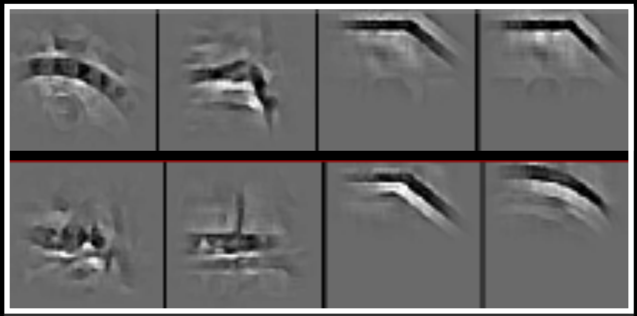


Sum

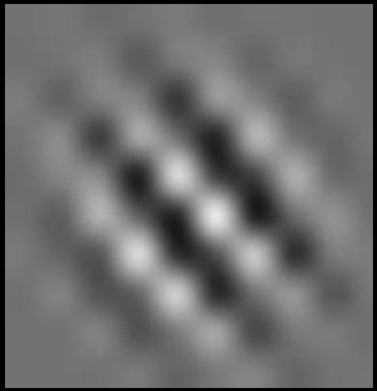


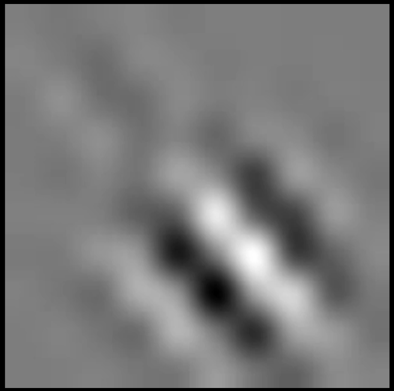
Role of Pooling

- Spatial pooling
  - Invariance to small transformations
  - Larger receptive fields (see more of input)



Visualization technique from [Le et al. NIPS'10]:



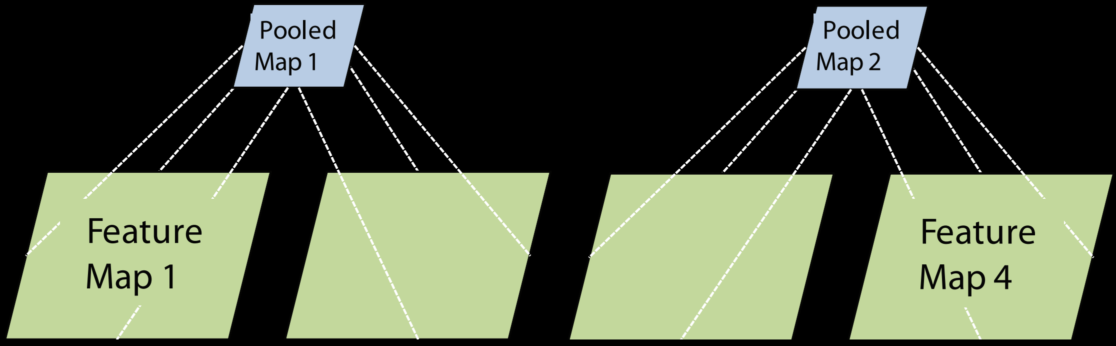


Zeiler, Taylor, Fergus [ICCV 2011]

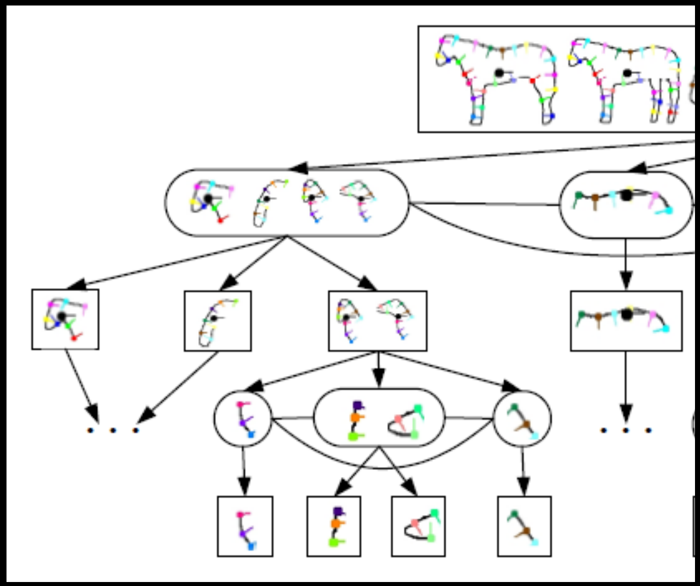
Videos from: http://ai.stanford.edu/~quocle/TCNNweb

Role of Pooling

- Pooling across feature groups
  - Additional form of inter-feature competition
  - Gives AND/OR type behavior via (sum / max)
  - Compositional models of Zhu, Yuille



[Zeiler et al., '11]



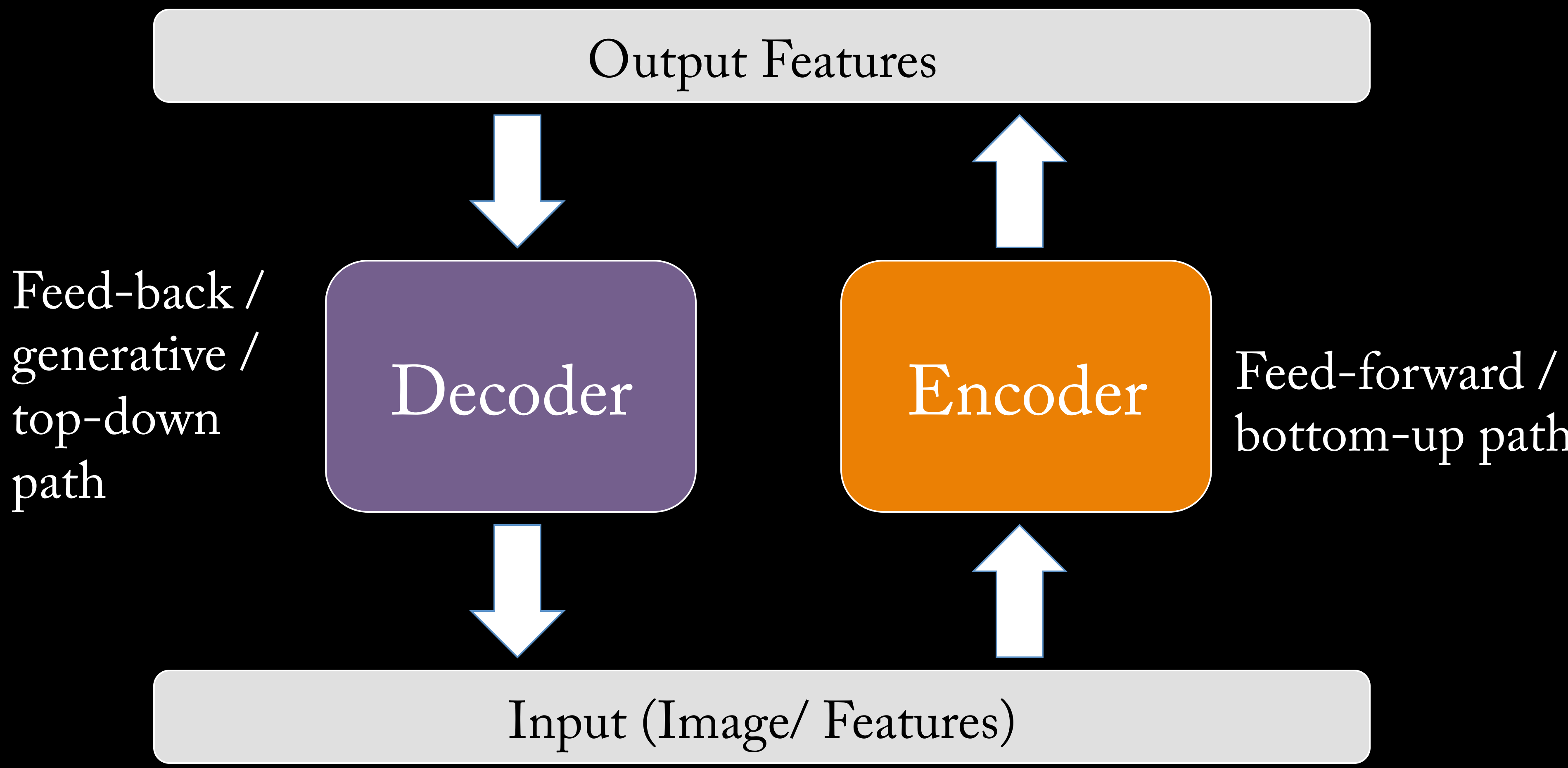
Chen, Zhu, Lin, Yuille, Zhang [NIPS 2007]

Unsupervised Learning

- Only have class labels at top layer
- Intermediate layers have to be trained unsupervised

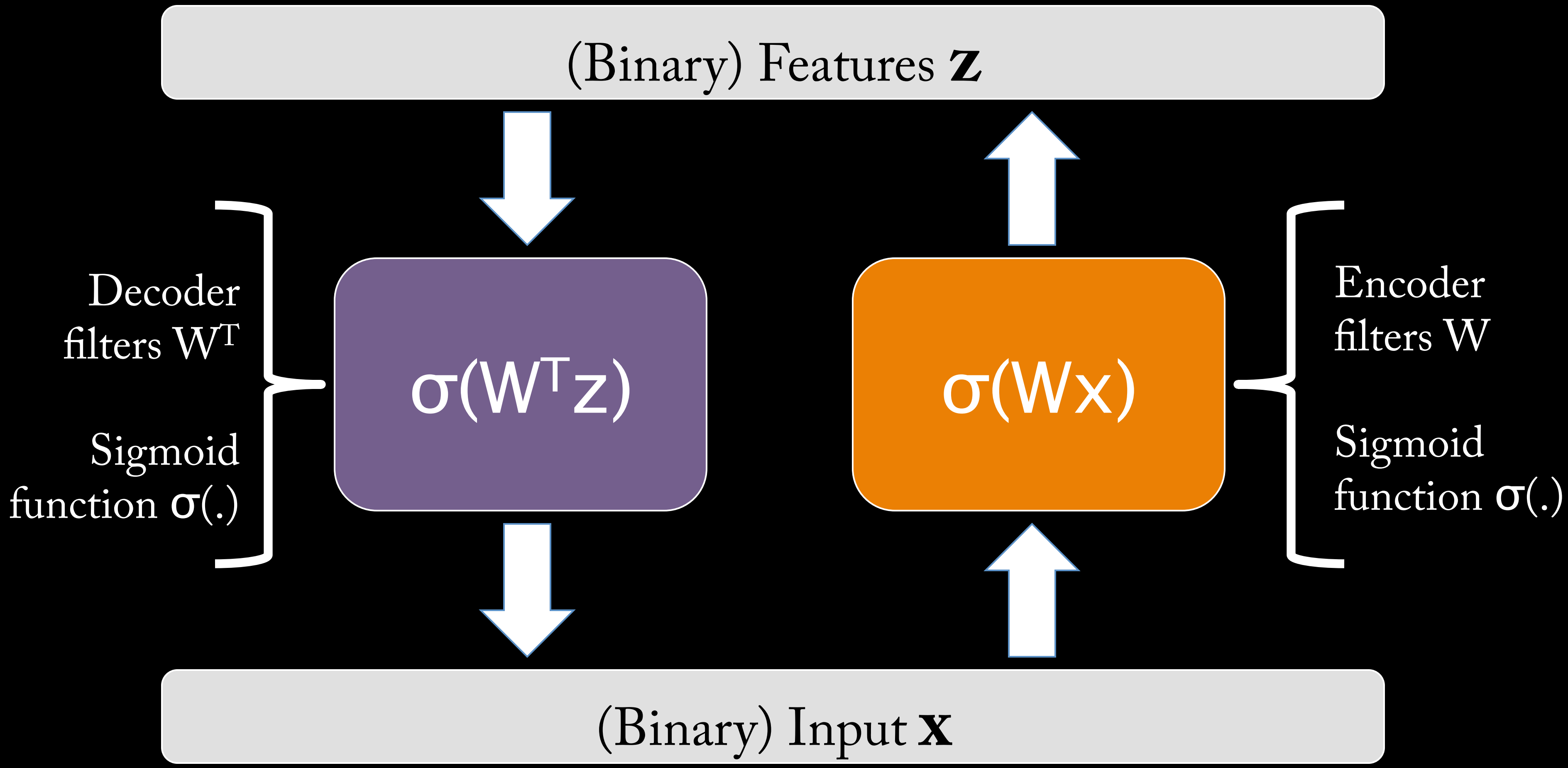
- Reconstruct input
  - 1st layer: image
  - Subsequent layers: features from layer beneath
  - Need constraint to avoid learning identity

Auto-Encoder



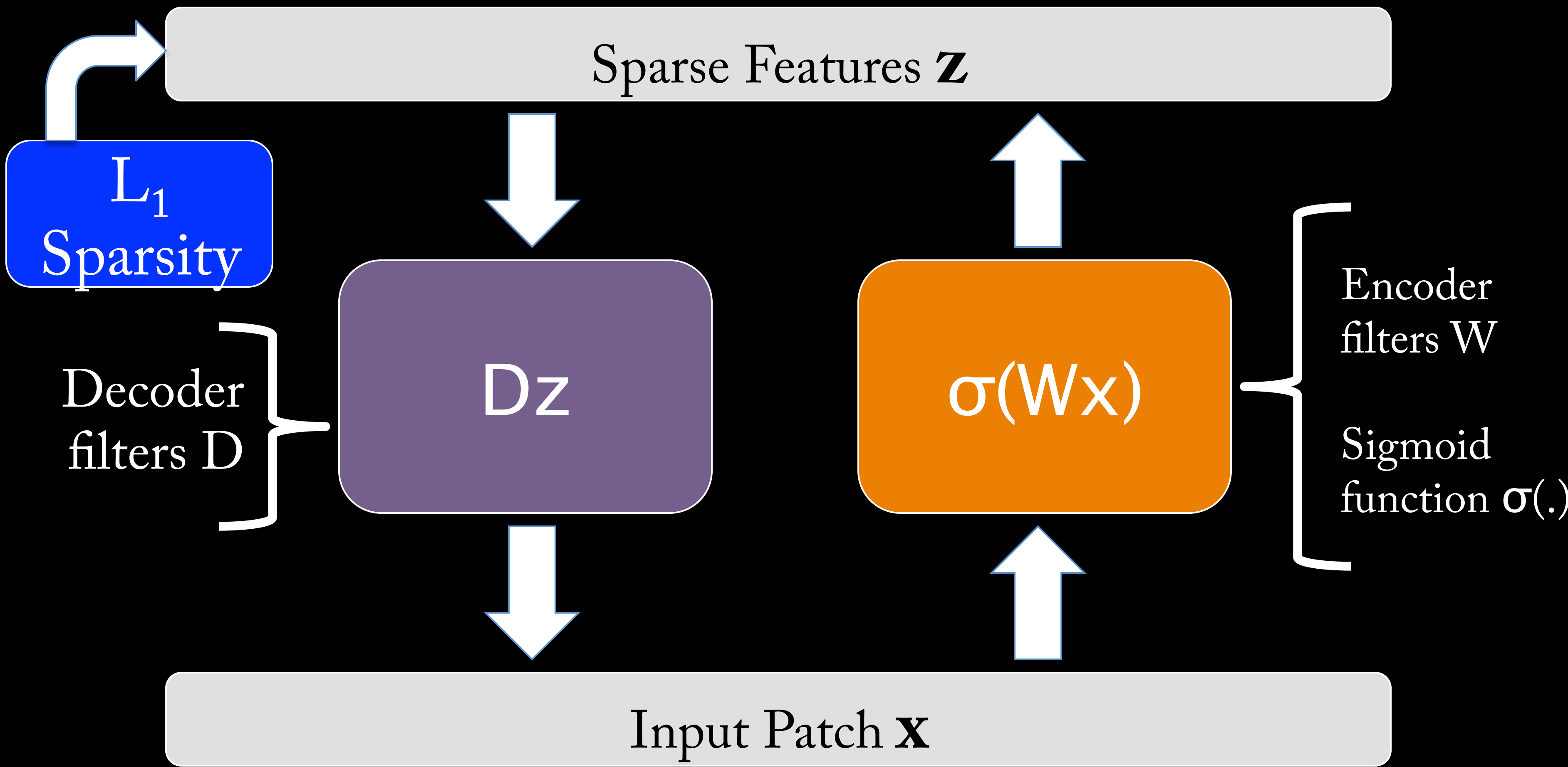
Auto-Encoder Example 1

- Restricted Boltzmann Machine [Hinton '02]



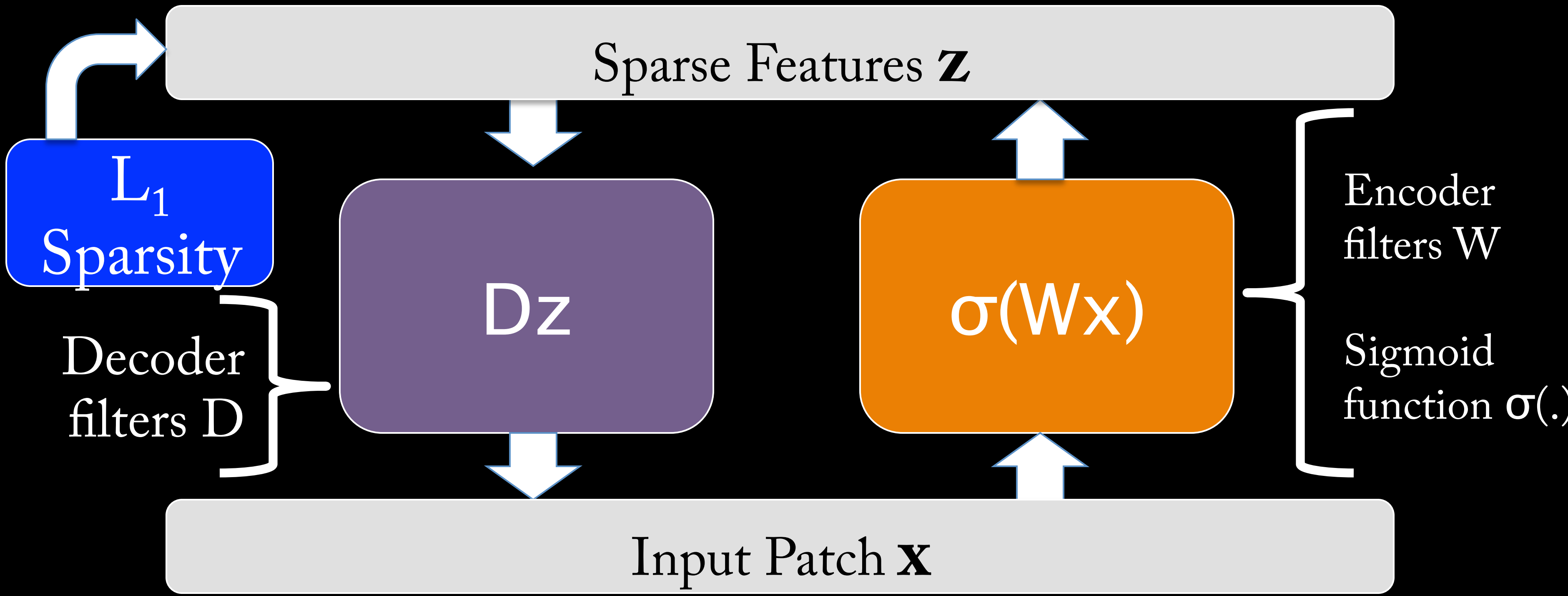
Auto-Encoder Example 2

- Predictive Sparse Decomposition [Ranzato et al., '07]



Auto-Encoder Example 2

- Predictive Sparse Decomposition [Kavukcuoglu et al., '09]



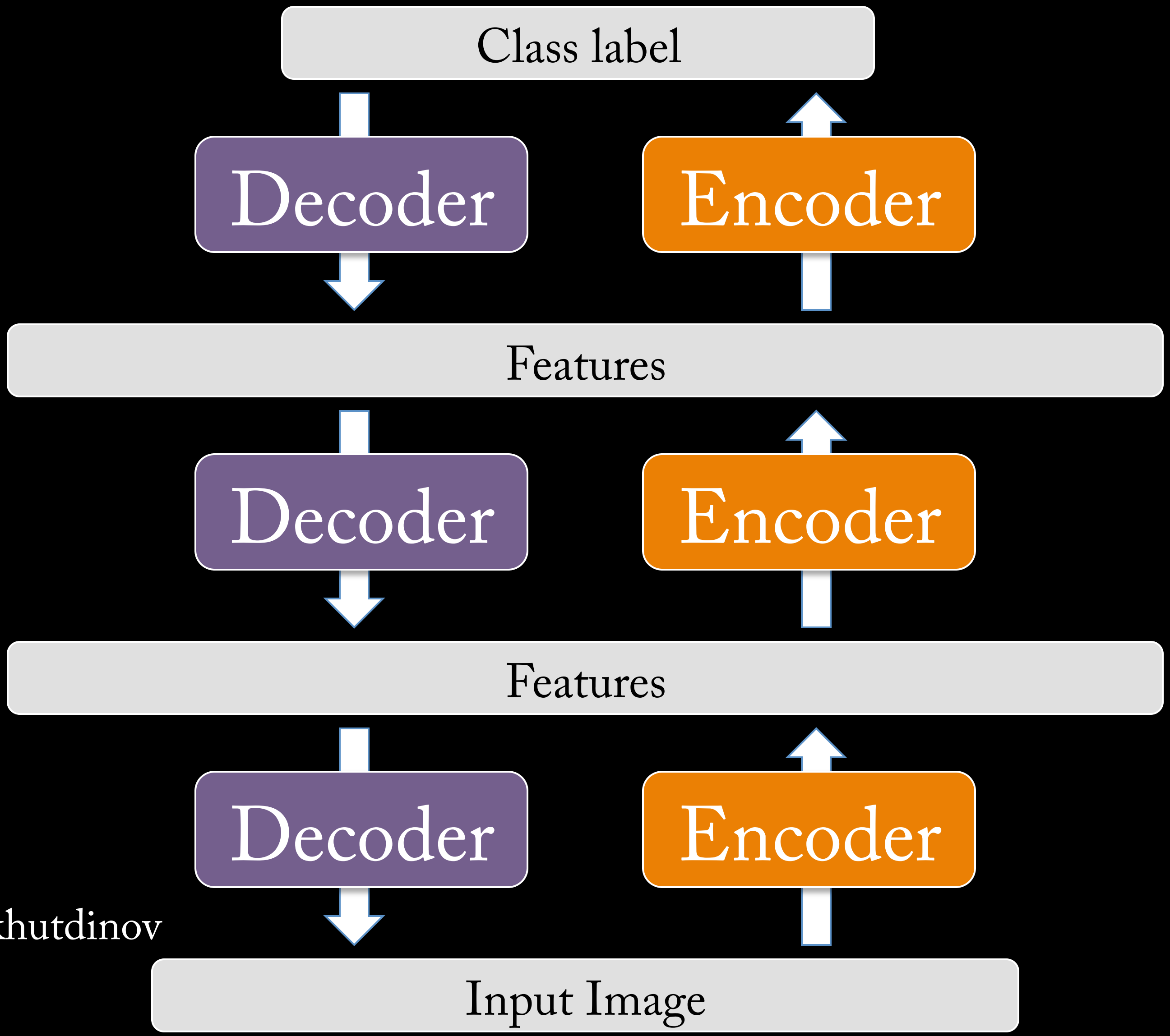
$$\min_{D,W,z} \underbrace{\|Dz - x\|_2^2 + \lambda|z|_1}_{Decoder} + \underbrace{\|\sigma(Wx) - z\|_2^2}_{Encoder}$$

Training

Taxonomy of Approaches

- Autoencoder (most Deep Learning methods)
  - RBMs / DBMs
  - Denoising autoencoders
  - Predictive sparse decomposition
- Decoder-only
  - Sparse coding
  - Deconvolutional Nets
- Encoder-only
  - Neural nets (supervised)

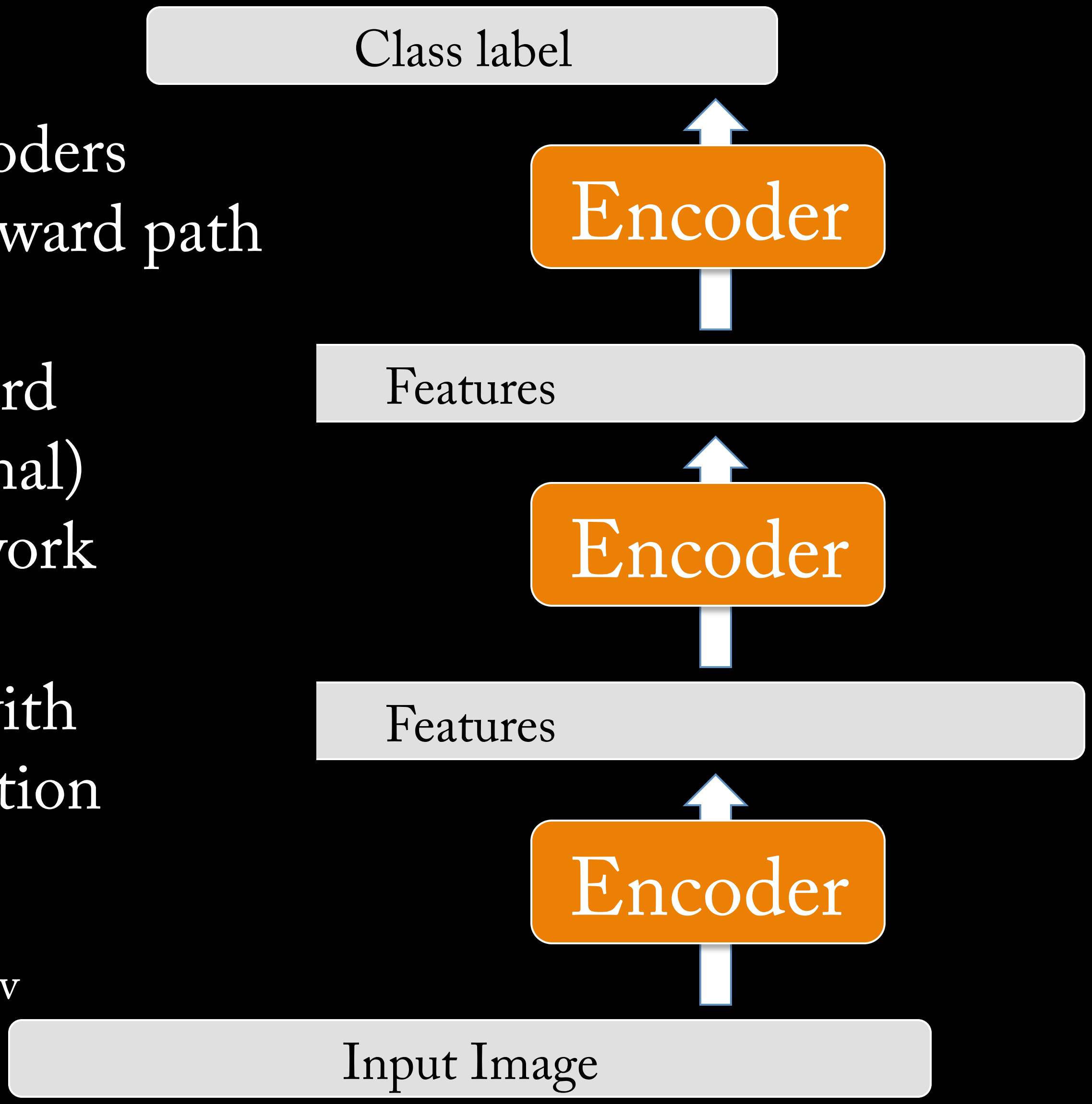
Stacked Auto-Encoders



[Hinton & Salakhutdinov Science '06]

At Test Time

- Remove decoders
- Use feed-forward path
- Gives standard (Convolutional) Neural Network
- Can refine with backpropagation



[Hinton & Salakhutdinov Science '06]

Semi-Supervised Training (2 phases)

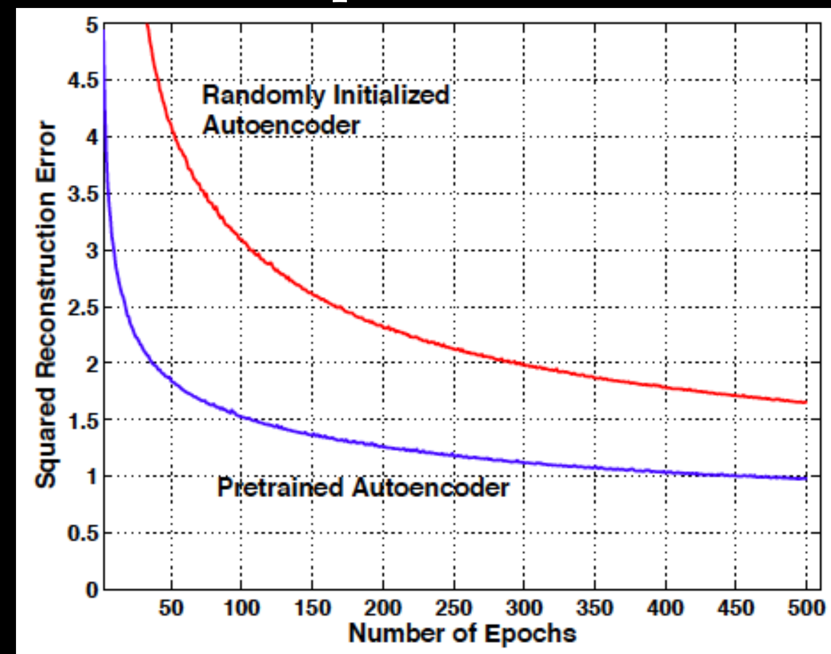
1. Unsupervised pre-training
   - Get parameters into right ball-park

2. Then supervised refinement (backpropagation)
   - Find local optima

- Helps to avoid local minima
  - Highly non-convex cost

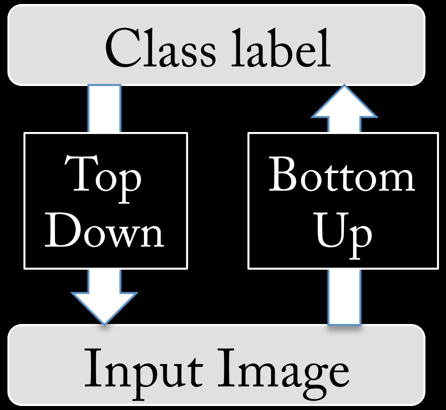
- Most common training paradigm in Deep Learning

Shallow Network / Deep Network



[Hinton & Salakhutdinov, Science '06]

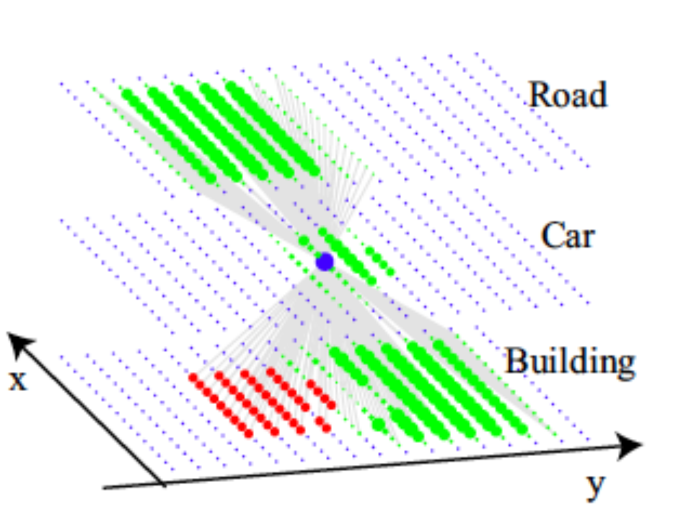
Information Flow in Vision Models



- Top-down (TD) vs bottom-up (BU)

- In Vision typically:
  BU appearance + TD shape
  – Example 1: MRF's
  – Example 2: Parts & Structure models

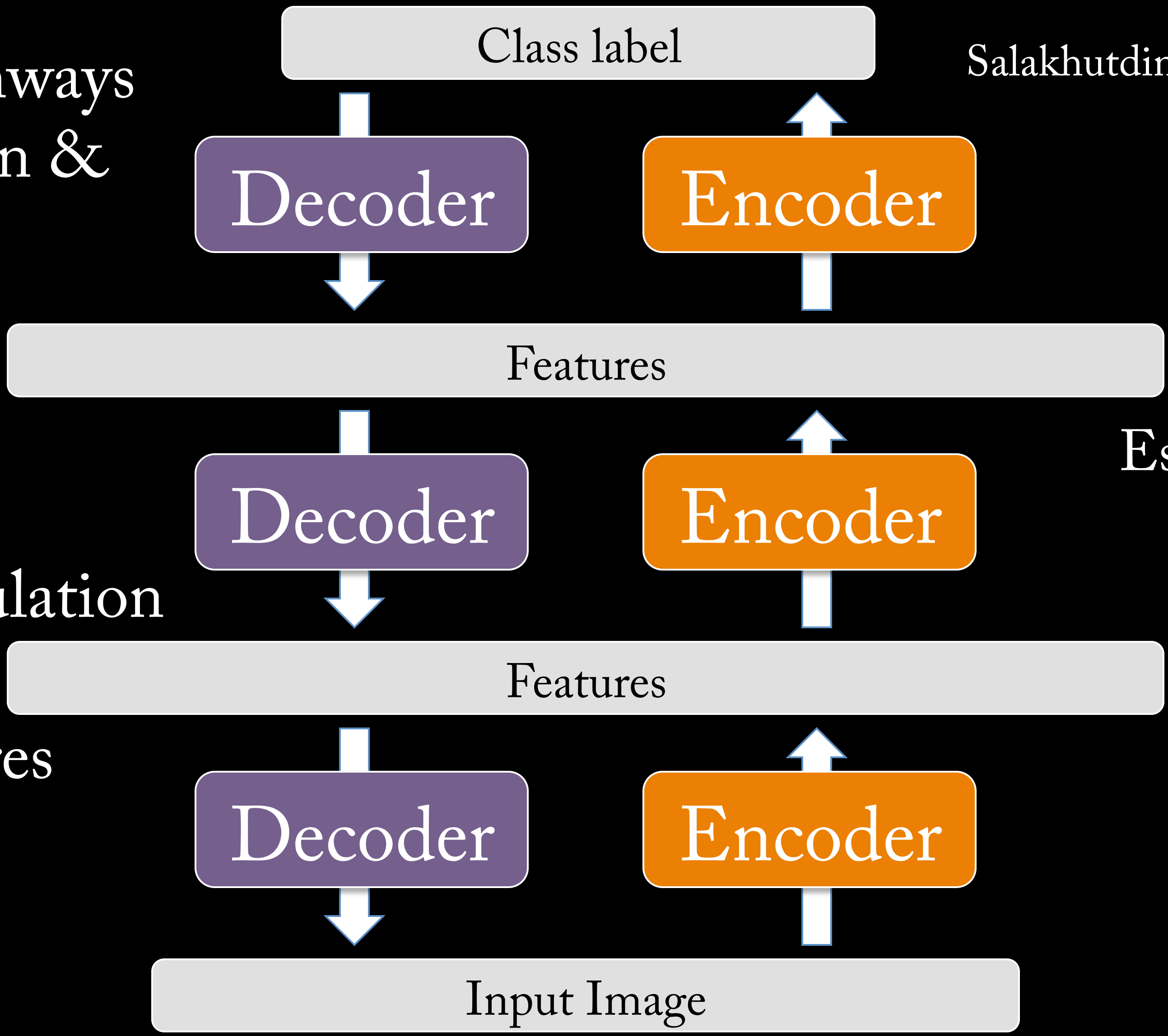
- TD context models
  – E.g. Torralba et al. NIPS'05



a) Incoming messages to a car node.

Deep Boltzmann Machines

Both pathways use at train & test time



Salakhutdinov & Hinton AISTATS'09

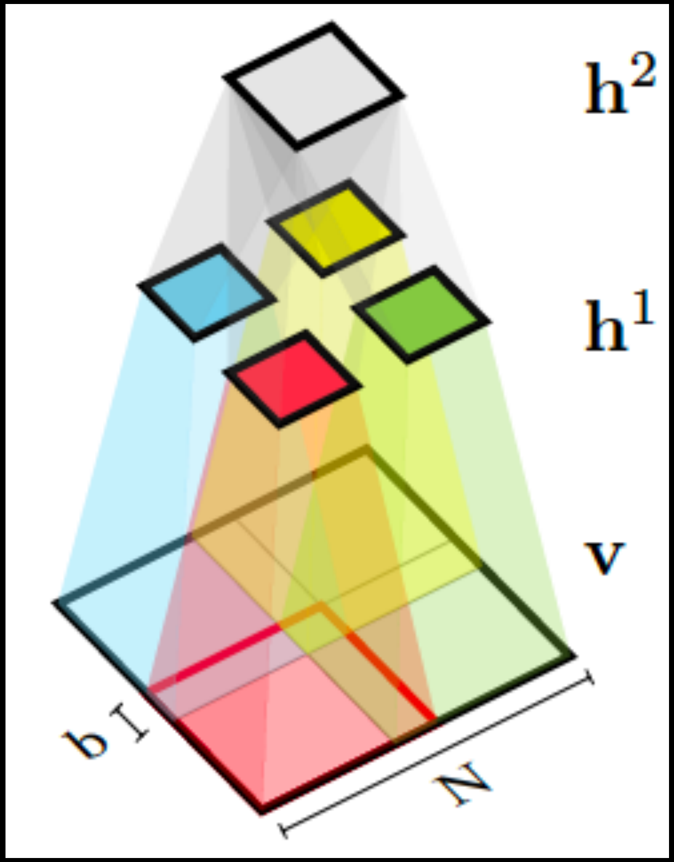
See also:

Eslami et al. CVPR'12

TD modulation of BU features

Deep Boltzmann Machines

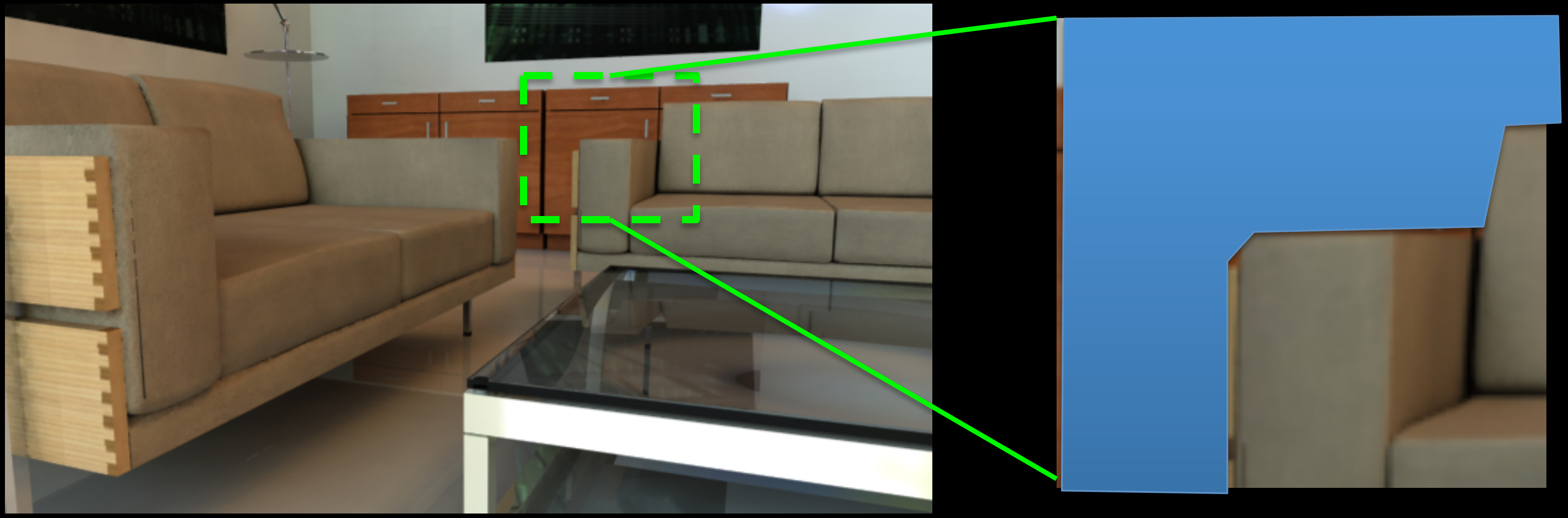
- Shape Boltzmann Machine
  - Eslami et al. CVPR'12
- 2 Hidden layers
  - Layer 1: tiled
  - Layer 2: densely connected
- Joint training of all layers
  - Only layer 2 can see whole image
  - Layer 2 crucial for training layer 1



Model samples for fixed h^2

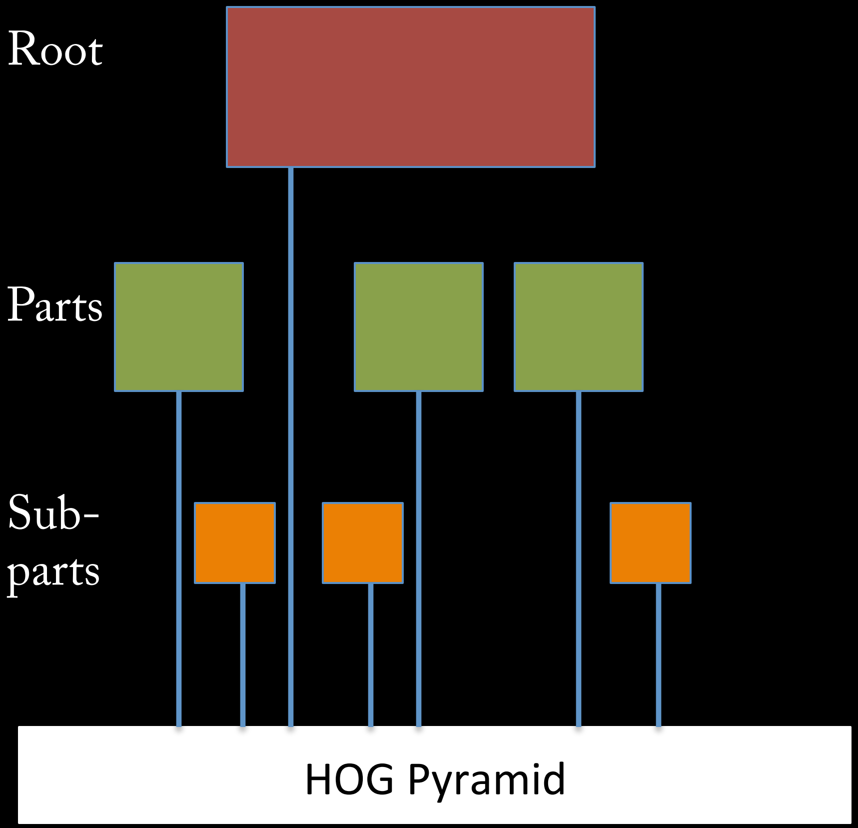
Why is Top-Down important?

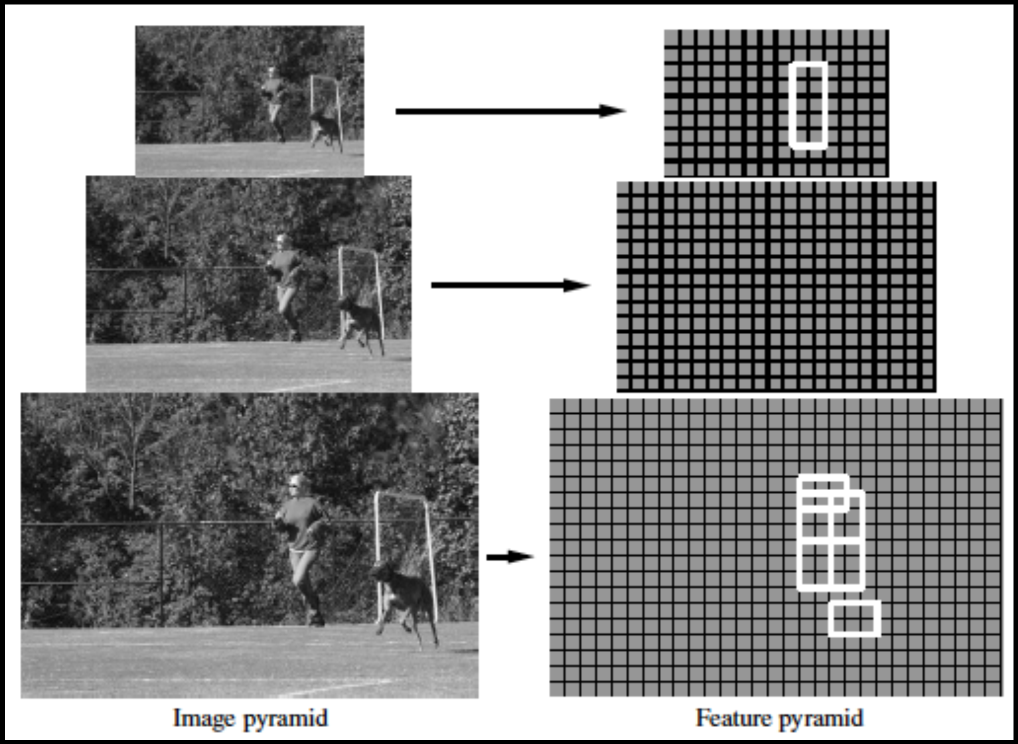
- Example: Occlusion
- BU alone can't separate sofa from cabinet
- Need TD information to focus on relevant part of region



Multi-Scale Models

- E.g. Deformable Parts Model
  - [Felzenszwalb et al. PAMI'10], [Zhu et al. CVPR'10]
  - Note: Shape part is hierarchical

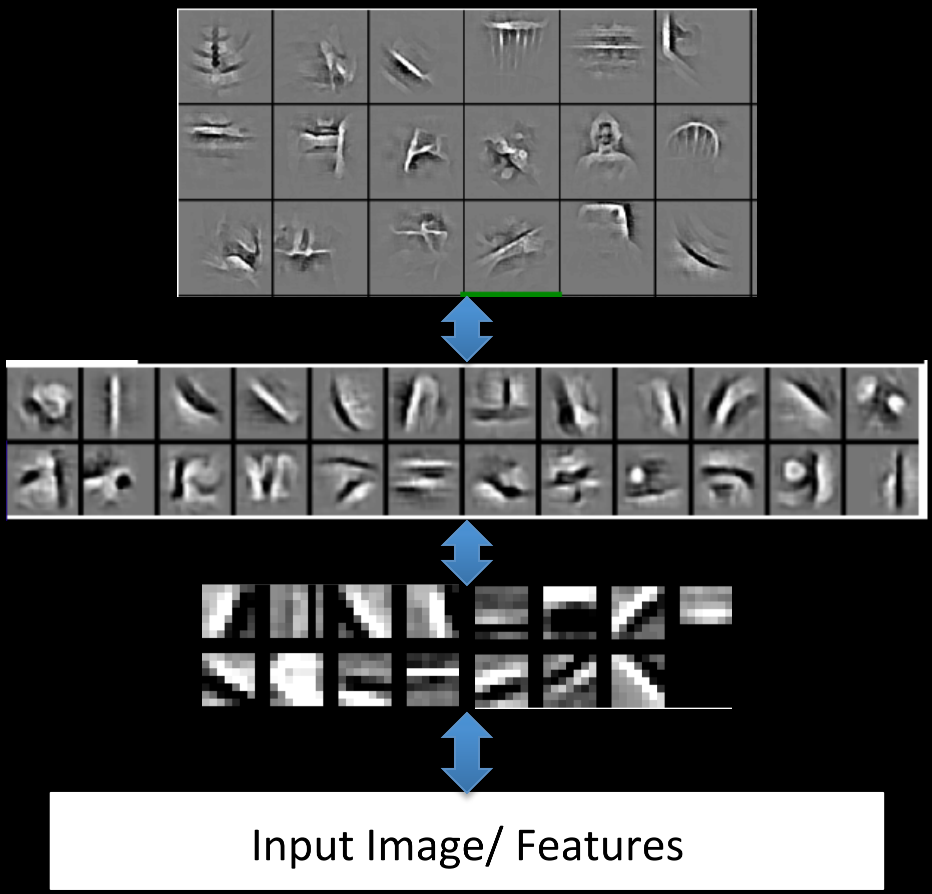




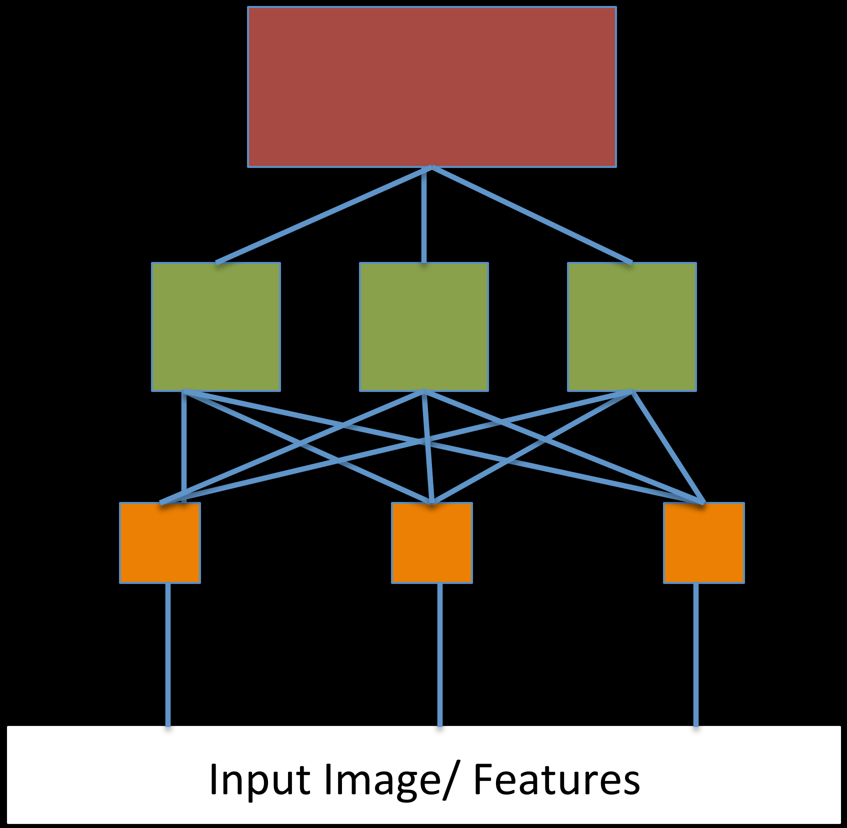
[Felzenszwalb et al. PAMI'10]

Hierarchical Model

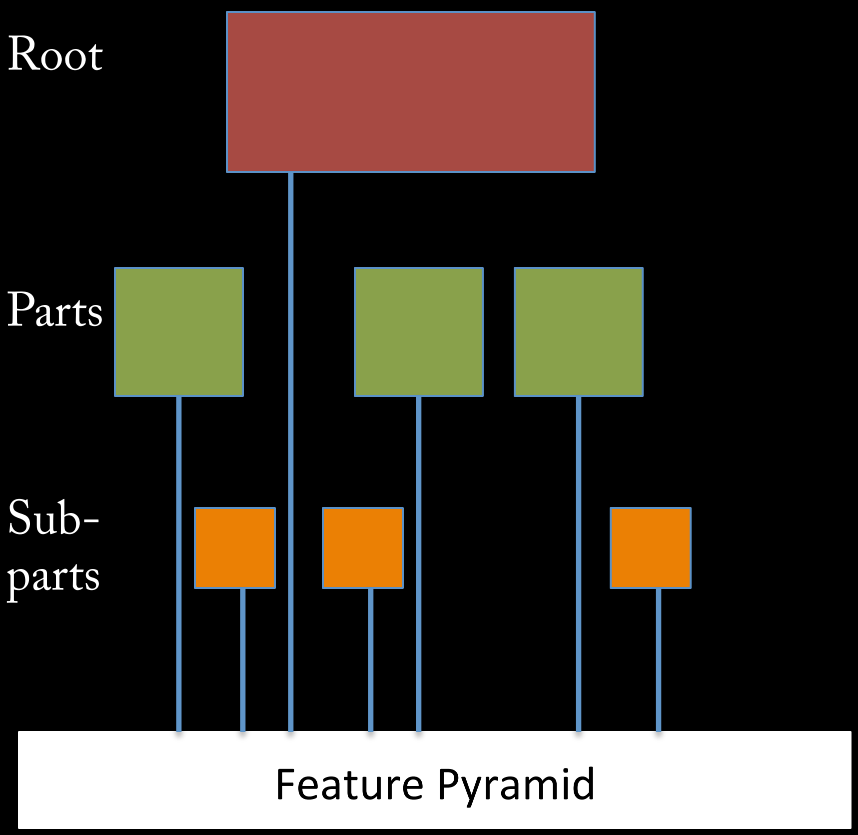
- Most Deep Learning models are hierarchical



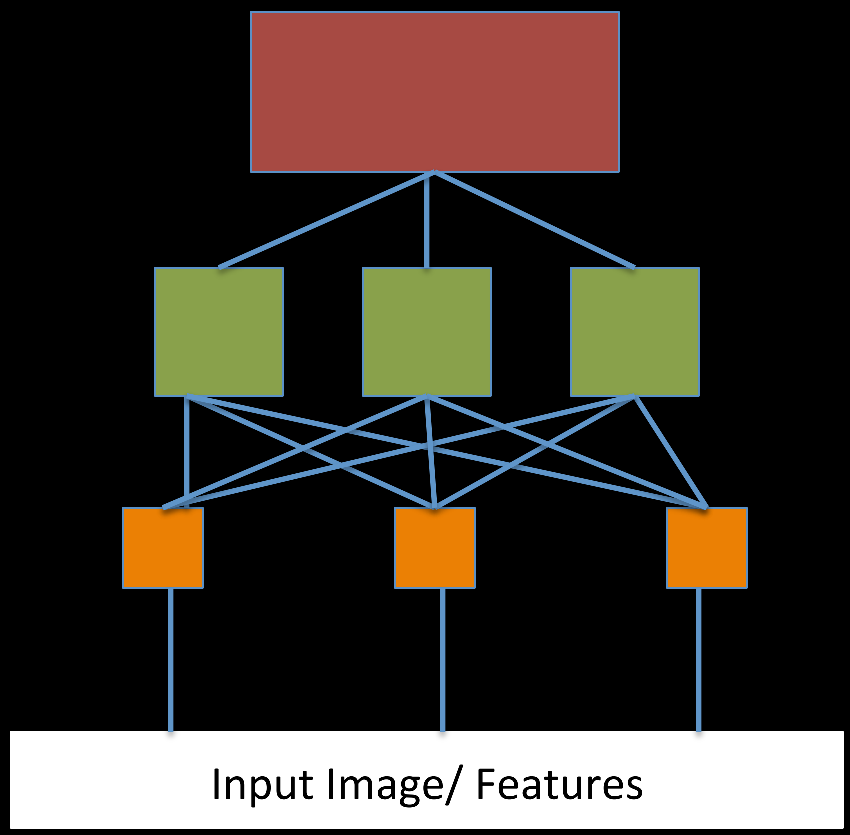
[Zeiler et al. ICCV'11]



Multi-scale vs Hierarchical

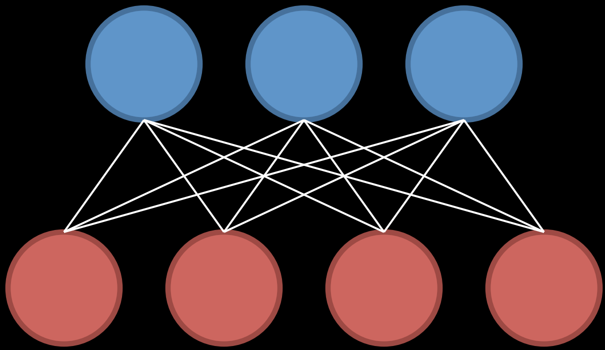


Appearance term of each part is independent of others

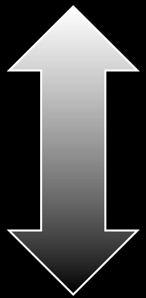


Parts at one layer of hierarchy depend on others

Structure Spectrum



- Learn everything
  - Homogenous architecture
  - Same for all modalities
  - Only concession topology (2D vs 1D)

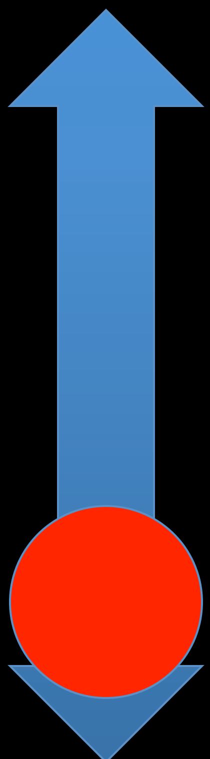


How much learning?

- Build vision knowledge into structure
  - Shape, occlusion etc.
  - Stochastic grammars, parts and structure models

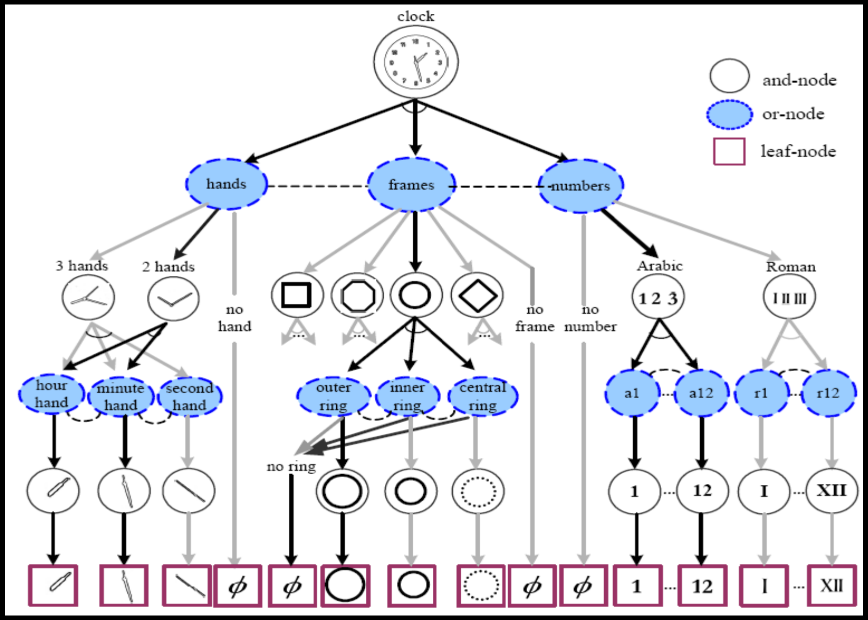
Structure Spectrum

Learn



Hand specify

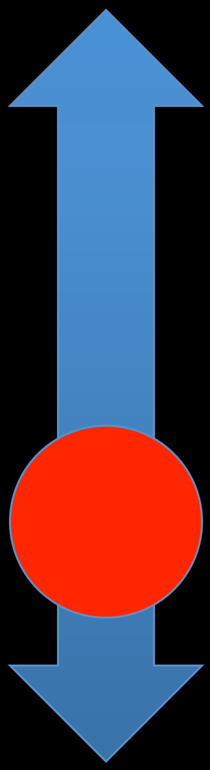
- Stochastic Grammar Models
  - Set of production rules for objects
  - Zhu & Mumford, Stochastic Grammar of Images, F&T 2006



[S.C. Zhu et al.]

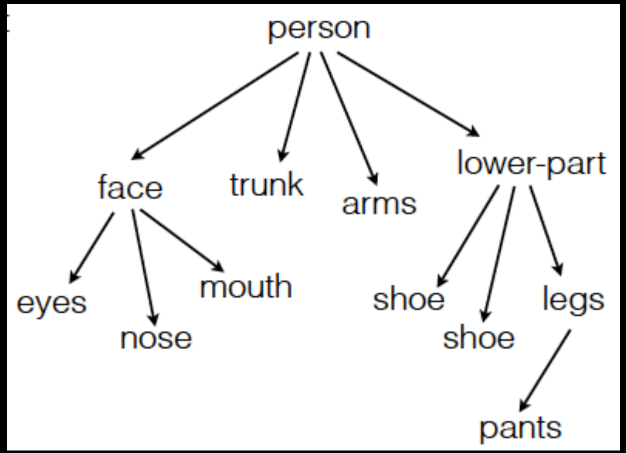
Structure Spectrum

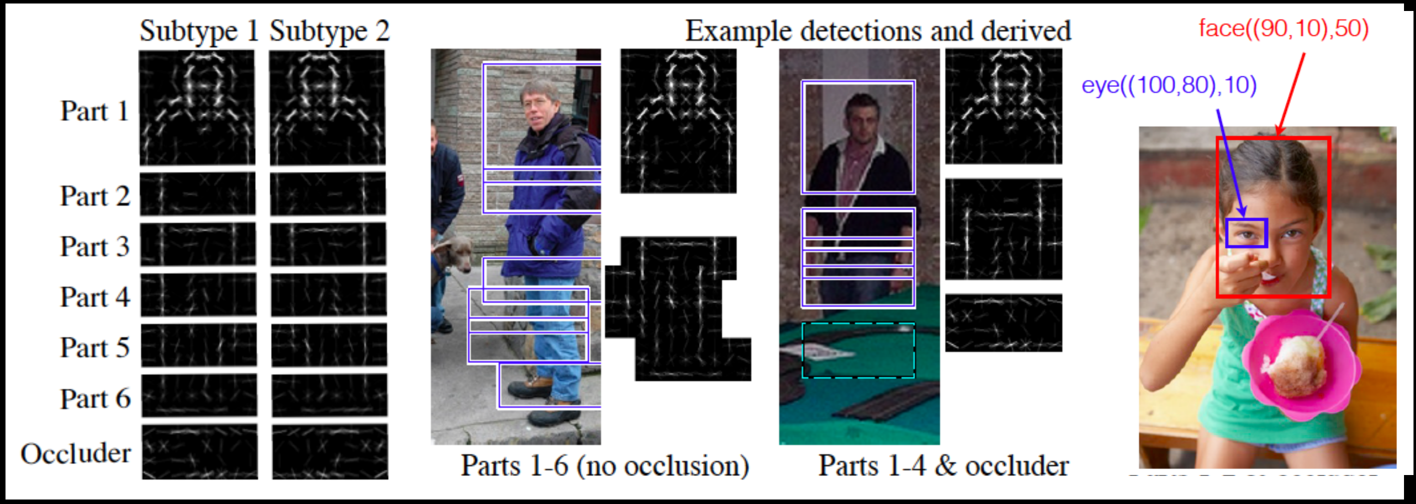
Learn



Hand specify

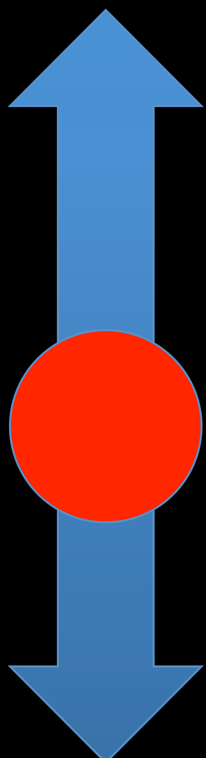
- R. Girshick, P. Felzenszwalb, D. McAllester, Object Detection with Grammar Models, NIPS 2011
- Learn local appearance & shape





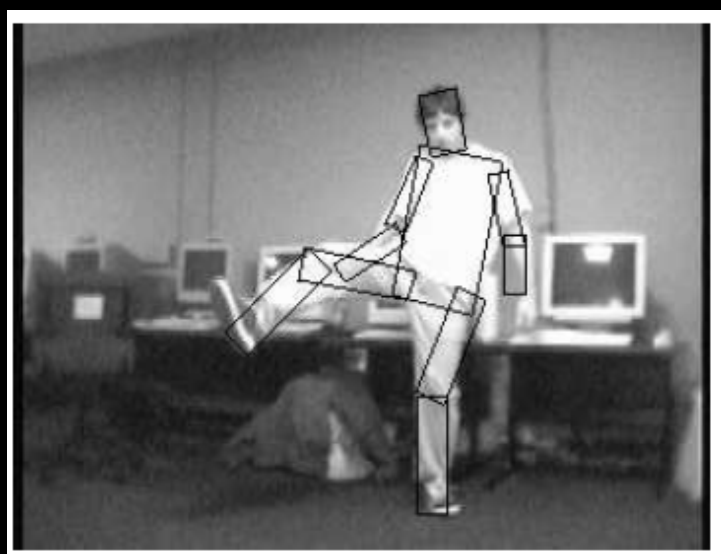
Structure Spectrum

Learn

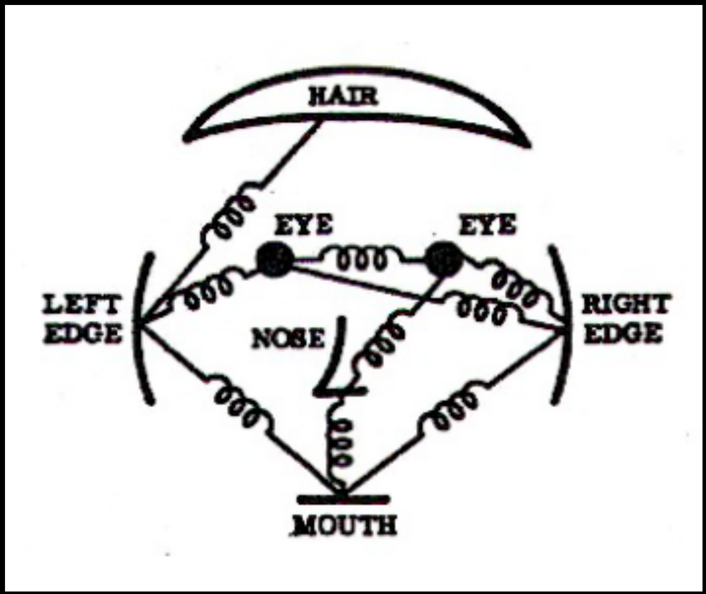


Hand specify

- Parts and Structure models
  - Defined connectivity graph
  - Learn appearance / relative position



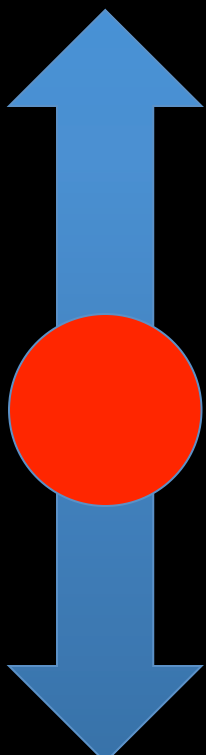
[Felzenszwalb & Huttenlocher CVPR'00]



[Fischler and R. Elschlager 1973]

Structure Spectrum

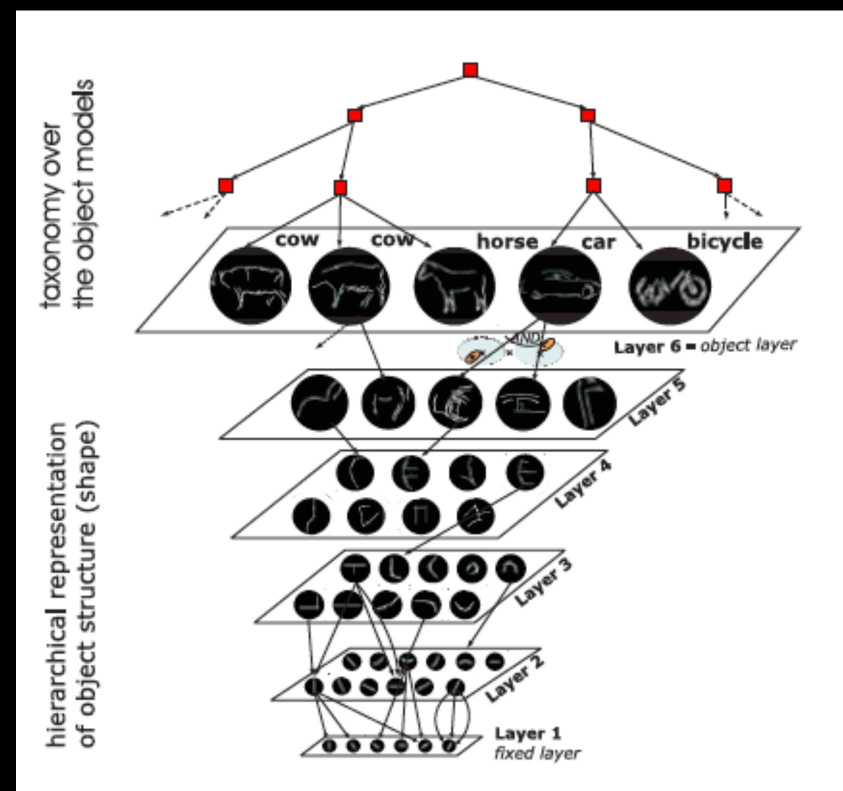
Learn



Hand specify

- Fidler et al. ECCV'10
- Fidler & Leonardis CVPR'07

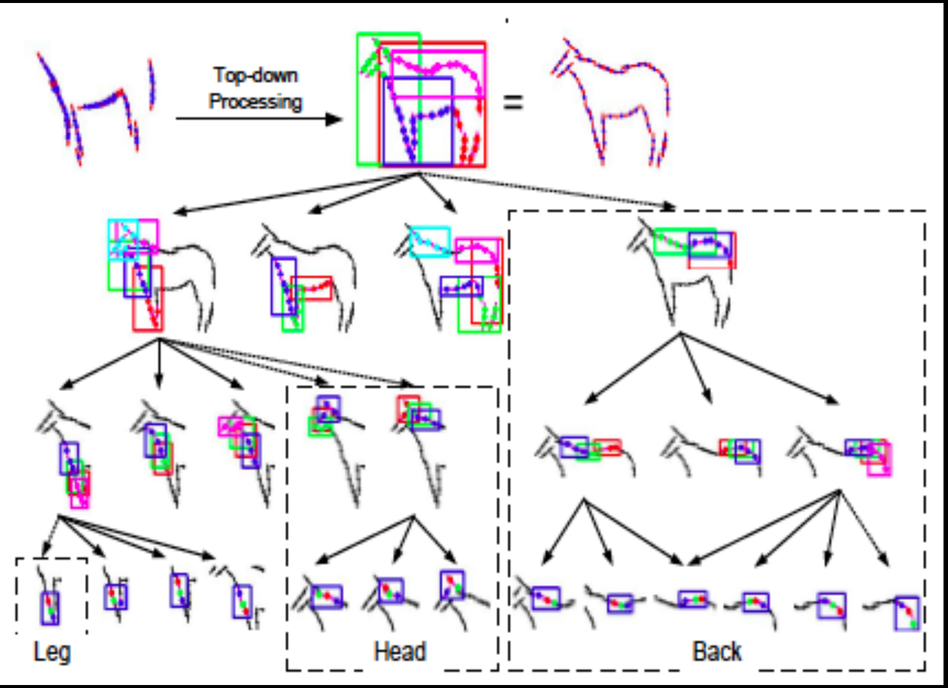
- Hierarchy of parts and structure models



Structure Spectrum

Learn

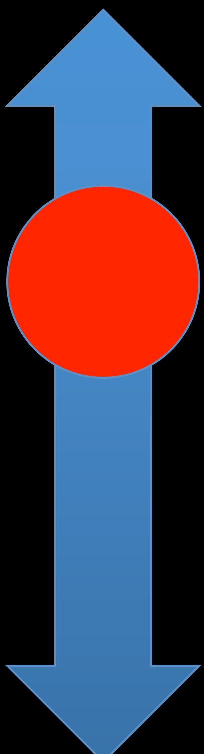
- Leo Zhu, Yuanhao Chen, Alan Yuille & collaborators
  - Recursive composition, AND/OR graph
  - Learn # units at layer



Hand specify

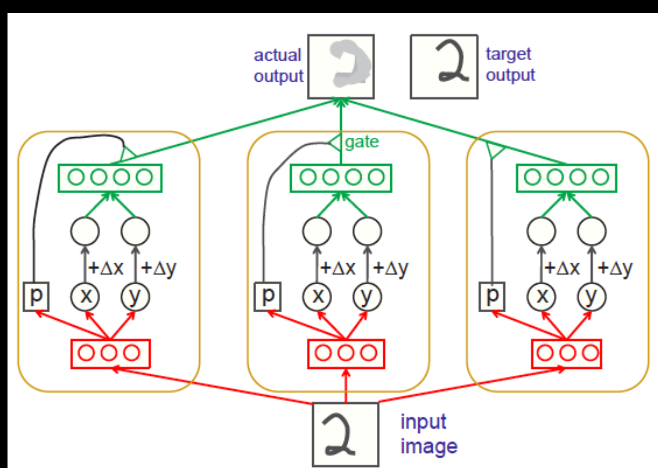
Structure Spectrum

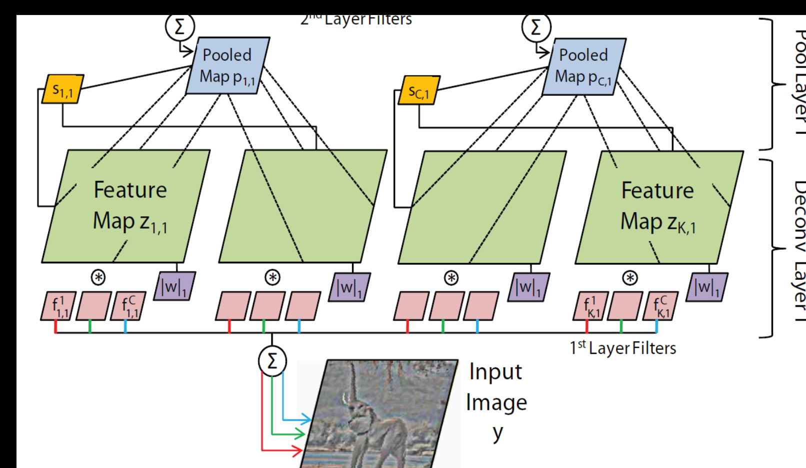
Learn



Hand specify

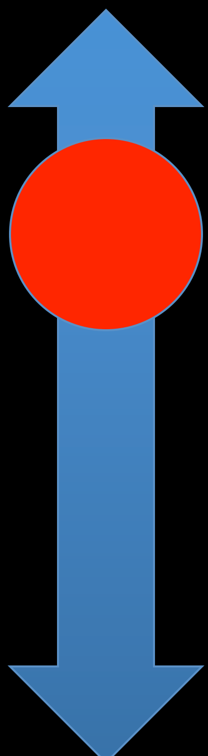
- Transforming Auto-Encoders
  - [Hinton et al. ICANN'11]
- Deconvolutional Networks
  - [Zeiler et al. ICCV'11]
- Explicit representation of what/where





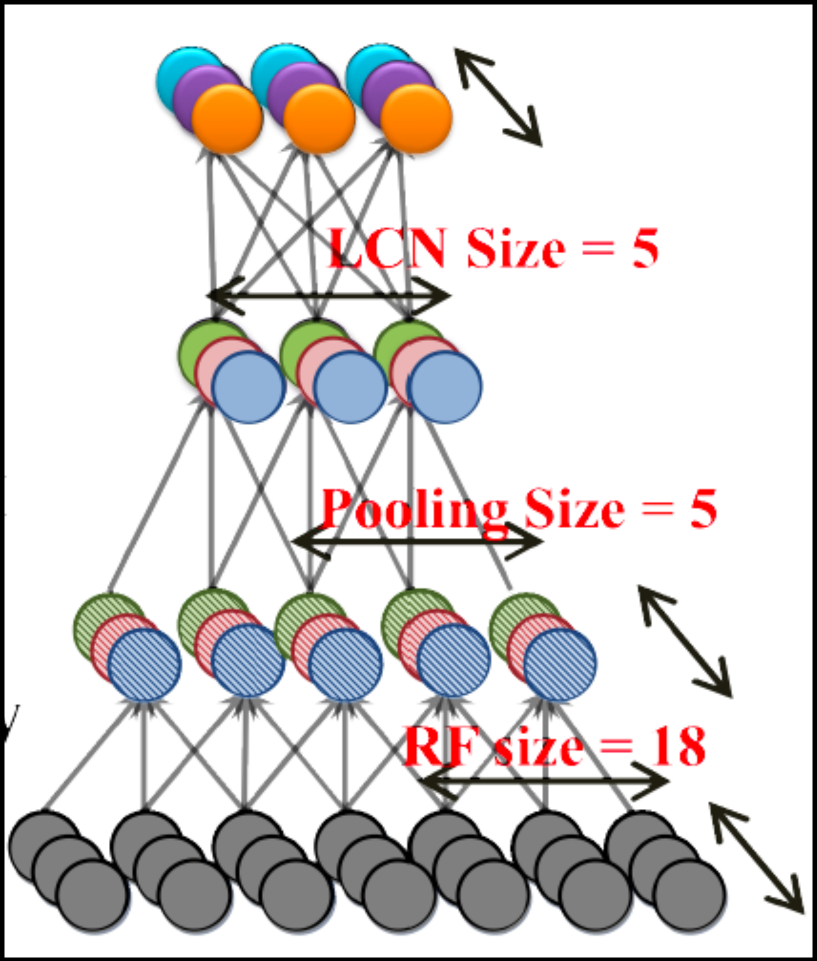
Structure Spectrum

Learn



Hand specify

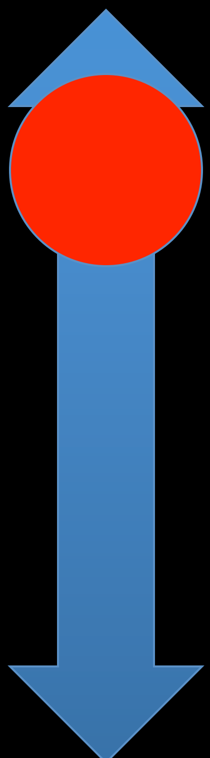
- Neural Nets / Auto-encoders
  - Dedicated pooling / LCN layers
  - No separation of what/where
  - Modality independent (e.g. speech, images)



[Le et al., ICML'12]

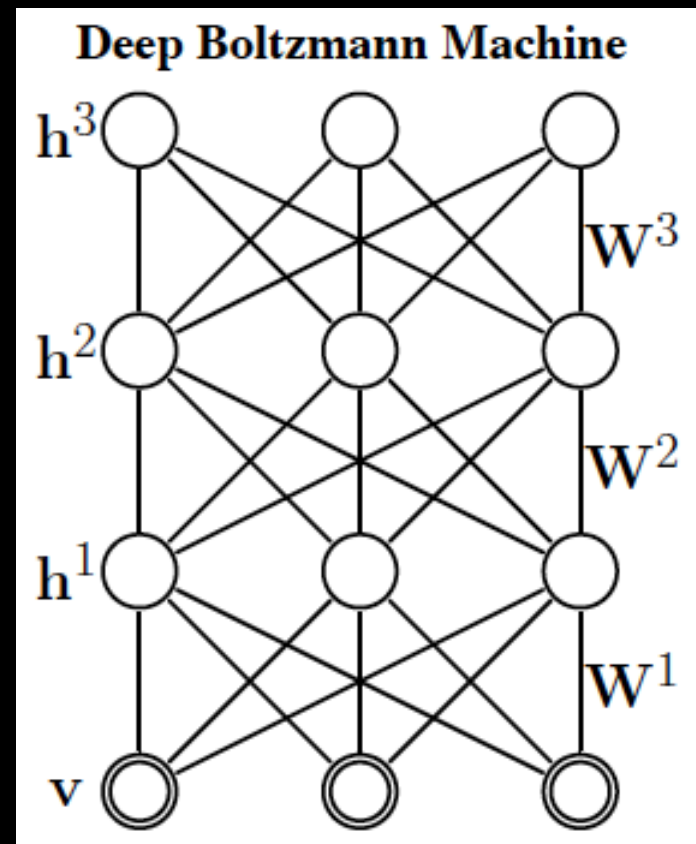
Structure Spectrum

Learn



Hand specify

- Boltzmann Machines
  - Homogenous architecture
  - No separation of what/where
  - Modality independent (e.g. speech, images)



[Salakhutdinov & Hinton AISTATS'09]

Performance of Deep Learning

- State-of-the-art on some (simpler) datasets
- Classification
  - ILSVRC 2010 (~1.4M images)
    - NEC/UIUC Winners (Sparse coding)
  - Full ImageNet (~16M images @ 2011)
    - Le et al. ICML'12 15.8% (vs 9.3% Weston et al.)
- Video
  - Holywood 2 (Action Recognition): Le et al. CVPR'11 53.3% (vs 50.9%)
- Detection
  - INRIA Pedestrians: Sermanet & LeCun (6.6% vs 8.6% miss rate @ 1FPPI)
- Not yet state-of-the-art on more challenging ones (e.g. PASCAL VOC Detection)

NIPS 2012: The Return of Convnets

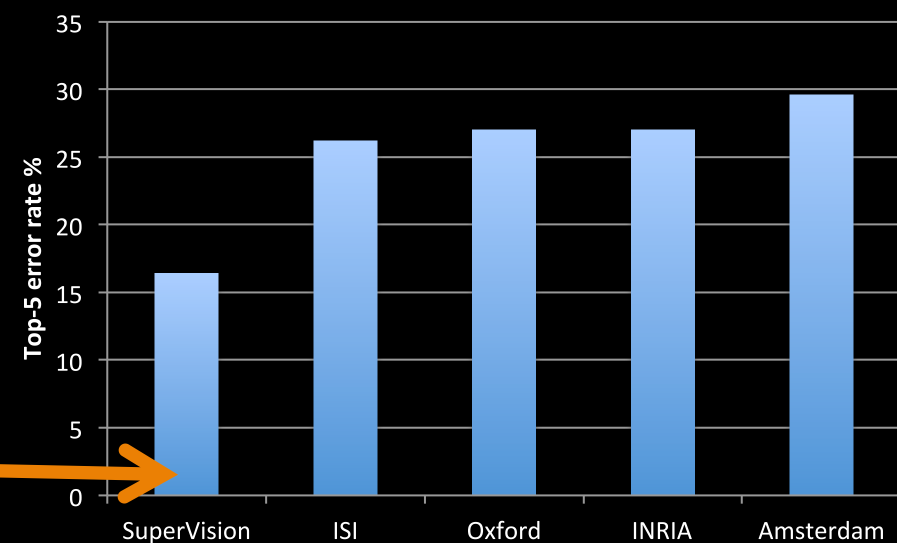
ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

ImageNet 2012 classification competition results



Summary

- Unsupervised Learning of Feature Hierarchies
  - Detailed explanation in following talks

- Showing promise on vision benchmarks
- Success in other modalities (speech, text)

- But few Deep Learning papers at CVPR!

Deep Learning & Feature Learning Methods for Vision

CVPR 2012 Tutorial

Rob Fergus (NYU)
Kai Yu (Baidu)
Marc' Aurelio Ranzato (Google)
Honglak Lee (Michigan)
Ruslan Salakhutdinov (U. Toronto)
Graham Taylor (University of Guelph)

Further Resources

- CVPR 2012 tutorial on Deep Learning http://cs.nyu.edu/~fergus/tutorials/deep_learning_cvpr12/
- http://deeplearning.net/
- http://www.cs.toronto.edu/~hinton/csc2515/deeprefs.html
- NIPS 2011 workshop on Deep Learning and Unsupervised Feature Learning
  - http://deeplearningworkshopnips2011.wordpress.com/
- Torch5 http://torch5.sourceforge.net/

Exam Questions

- 1. In classical approaches to feature learning (e.g. ConvNets), learning was purely supervised. What form does learning take in recent Deep Learning approaches?

Exam Questions

- 2. Normalization is a key component in many Deep Learning approaches. What form does this take?

Exam Questions

- 3. In an auto-encoder, which of the following roles are performed by the decoder:

    (i) providing a feed-forward path for quick feature computation

    (ii) ensuring that the features reconstruct the input

    (iii) providing a target during training for the encoder

References

- [Slide 5]
- P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, Object Detection with Discriminatively Trained Part Based Models,IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 9, September 2010
- Zheng Song*, Qiang Chen*, Zhongyang Huang, Yang Hua, and Shuicheng Yan. Con-tex-tu-al-iz-ing Ob-ject De-tec-tion and Clas-si-fi-ca-tion. In CVPR'11. (* in-di-cates equal con-tri-bu-tion) [No. 1 per-for-mance in VOC'10 clas-si-fi-ca-tion task]
- [Slide 6]
- Finding the Weakest Link in Person Detectors, D. Parikh, and C. L. Zitnick, CVPR, 2011.
- [Slide 7]
- Gehler and Nowozin, On Feature Combination for Multiclass Object Classification, ICCV'09
- [Slide 8]
- http://www.amazon.com/Vision-David-Marr/dp/0716712849
- [Slide 10]
- Yoshua Bengio and Yann LeCun: Scaling learning algorithms towards AI, in Bottou, L. and Chapelle, O. and DeCoste, D. and Weston, J. (Eds), Large-Scale Kernel Machines, MIT Press, 2007

References

- [Slide 11]
- S. Lazebnik, C. Schmid, and J. Ponce, Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006
- [Slide 12]
- Christoph H. Lampert, Hannes Nickisch, Stefan Harmeling: "Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer", IEEE Computer Vision and Pattern Recognition (CVPR), Miami, FL, 2009
- [Slide 14] Riesenhuber, M. & Poggio, T. (1999). Hierarchical Models of Object Recognition in Cortex. Nature Neuroscience 2: 1019-1025.
- http://www.scholarpedia.org/article/Neocognitron
- K. Fukushima: "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", Biological Cybernetics, 36[4], pp. 193-202 (April 1980).
- Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE, 86(11):2278-2324, November 1998

References

- [Slide 30]
- Y-Lan Boureau, Jean Ponce, and Yann LeCun, A theoretical analysis of feature pooling in vision algorithms, Proc. International Conference on Machine learning (ICML'10), 2010
- [Slide 31]
- Q.V. Le, J. Ngiam, Z. Chen, D. Chia, P. Koh, A.Y. Ng , Tiled Convolutional Neural Networks. NIPS, 2010
- http://ai.stanford.edu/~quocle/TCNNweb
- Matthew D. Zeiler, Graham W. Taylor, and Rob Fergus, Adaptive Deconvolutional Networks for Mid and High Level Feature Learning, International Conference on Computer Vision(November 6-13, 2011)
- [Slide 32]
- Yuanhao Chen, Long Zhu, Chenxi Lin, Alan Yuille, Hongjiang Zhang. Rapid Inference on a Novel AND/OR graph for Object Detection, Segmentation and Parsing. NIPS 2007.

References

- [Slide 35]
- P. Smolensky, Parallel Distributed Processing: Volume 1: Foundations, D. E. Rumelhart, J. L. McClelland, Eds. (MIT Press, Cambridge, 1986), pp. 194–281.
- G. E. Hinton, Neural Comput. 14, 1711 (2002).
- [Slide 36]
- M. Ranzato, Y. Boureau, Y. LeCun. "Sparse Feature Learning for Deep Belief Networks". Advances in Neural Information Processing Systems 20 (NIPS 2007).
- [Slide 39]
- Hinton, G. E. and Salakhutdinov, R. R., Reducing the dimensionality of data with neural networks. Science, Vol. 313. no. 5786, pp. 504 - 507, 28 July 2006.
- [Slide 41]
- A. Torralba, K. P. Murphy and W. T. Freeman, Contextual Models for Object Detection using Boosted Random Fields, Adv. in Neural Information Processing Systems 17 (NIPS), pp. 1401-1408, 2005.

References

- [Slide 42]
- Ruslan Salakhutdinov and Geoffrey Hinton, Deep Boltzmann Machines, 12th International Conference on Artificial Intelligence and Statistics (2009).
- [Slide 44]
- P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, Object Detection with Discriminatively Trained Part Based Models,IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 9, September 2010
- Long Zhu, Yuanhao Chen, Alan Yuille, William Freeman. Latent Hierarchical Structural Learning for Object Detection. CVPR 2010.
- [Slide 45]
- Matthew D. Zeiler, Graham W. Taylor, and Rob Fergus, Adaptive Deconvolutional Networks for Mid and High Level Feature Learning, International Conference on Computer Vision(November 6-13, 2011)

References

- [Slide 48]
- S.C. Zhu and D. Mumford, A Stochastic Grammar of Images, Foundations and Trends in Computer Graphics and Vision, Vol.2, No.4, pp 259-362, 2006.
- [Slide 49]
- R. Girshick, P. Felzenszwalb, D. McAllester, Object Detection with Grammar Models, NIPS 2011
- [Slide 50]
- P. Felzenszwalb, D. Huttenlocher, Pictorial Structures for Object Recognition, International Journal of Computer Vision, Vol. 61, No. 1, January 2005
- M. Fischler and R. Elschlager. The Representation and Matching of Pictoral Structures. (1973)
- [Slide 51]
- S. Fidler, M. Boben, A. Leonardis. A coarse-to-fine Taxonomy of Constellations for Fast Multi-class Object Detection. ECCV 2010.
- S. Fidler and A. Leonardis. Towards Scalable Representations of Object Categories: Learning a Hierarchy of Parts. CVPR 2007.

References

- [Slide 52]
- Long Zhu, Chenxi Lin, Haoda Huang, Yuanhao Chen, Alan Yuille. Unsupervised Structure Learning: Hierarchical Recursive Composition, Suspicious Coincidence and Competitive Exclusion. ECCV 2008.
- [Slide 53]
- Hinton, G. E., Krizhevsky, A. and Wang, S, Transforming Auto-encoders. ICANN-11: International Conference on Artificial Neural Networks, 2011
- Matthew D. Zeiler, Graham W. Taylor, and Rob Fergus, Adaptive Deconvolutional Networks for Mid and High Level Feature Learning, International Conference on Computer Vision (November 6-13, 2011)
- [Slide 54]
- Q.V. Le, M.A. Ranzato, R. Monga, M. Devin, K. Chen, G.S. Corrado, J. Dean, A.Y. Ng., Building high-level features using large scale unsupervised learning. ICML, 2012.
- [Slide 55]
- Ruslan Salakhutdinov and Geoffrey Hinton, Deep Boltzmann Machines, 12th International Conference on Artificial Intelligence and Statistics (2009).

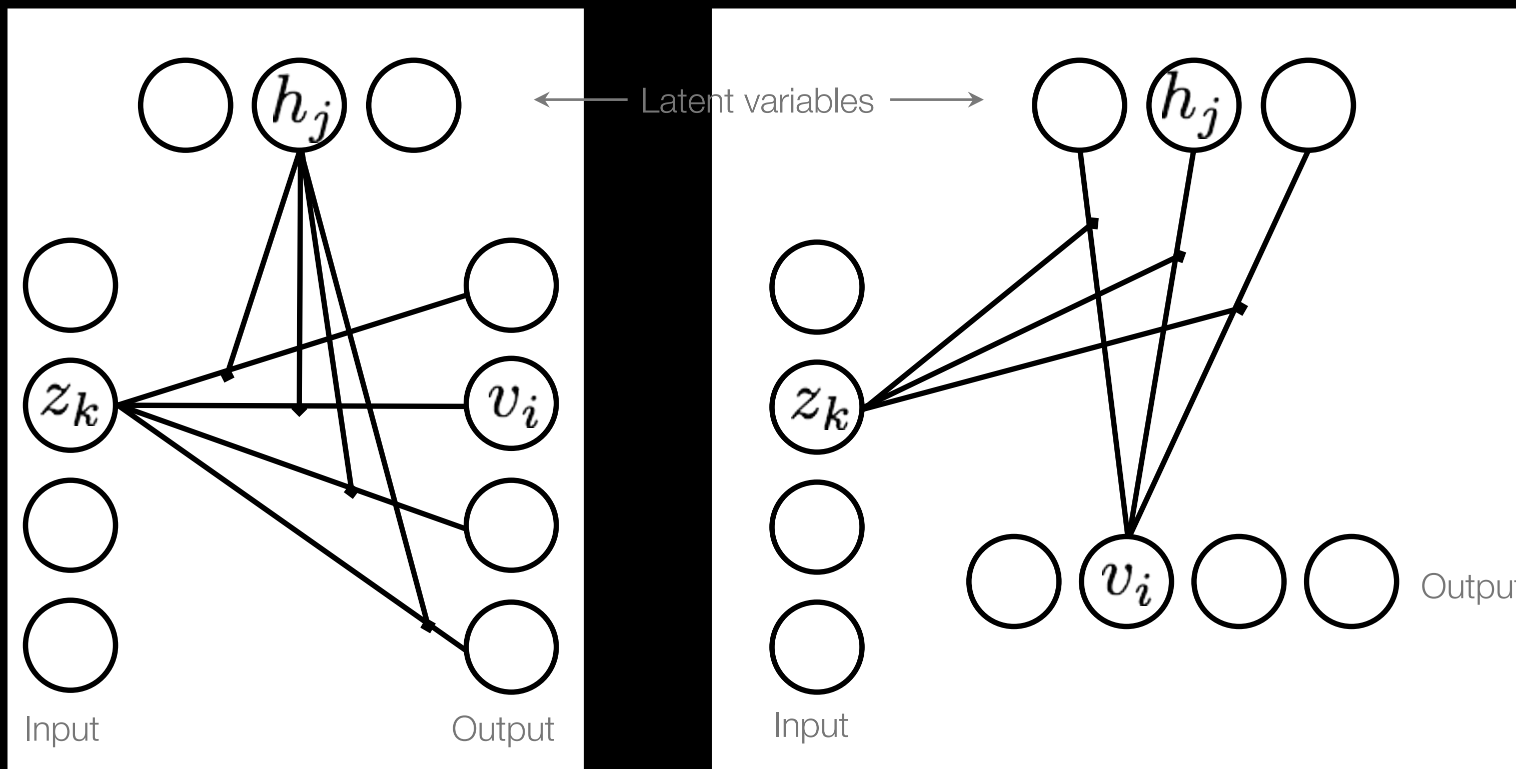
References

- [Slide 56]
- http://www.image-net.org/challenges/LSVRC/2010/
- Q.V. Le, M.A. Ranzato, R. Monga, M. Devin, K. Chen, G.S. Corrado, J. Dean, A.Y. Ng., Building high-level features using large scale unsupervised learning. ICML, 2012.
- Q.V. Le, W.Y. Zou, S.Y. Yeung, A.Y. Ng., Learning hierarchical spatio-temporal features for action recognition with independent subspace analysis, CVPR 2011

Application to Video

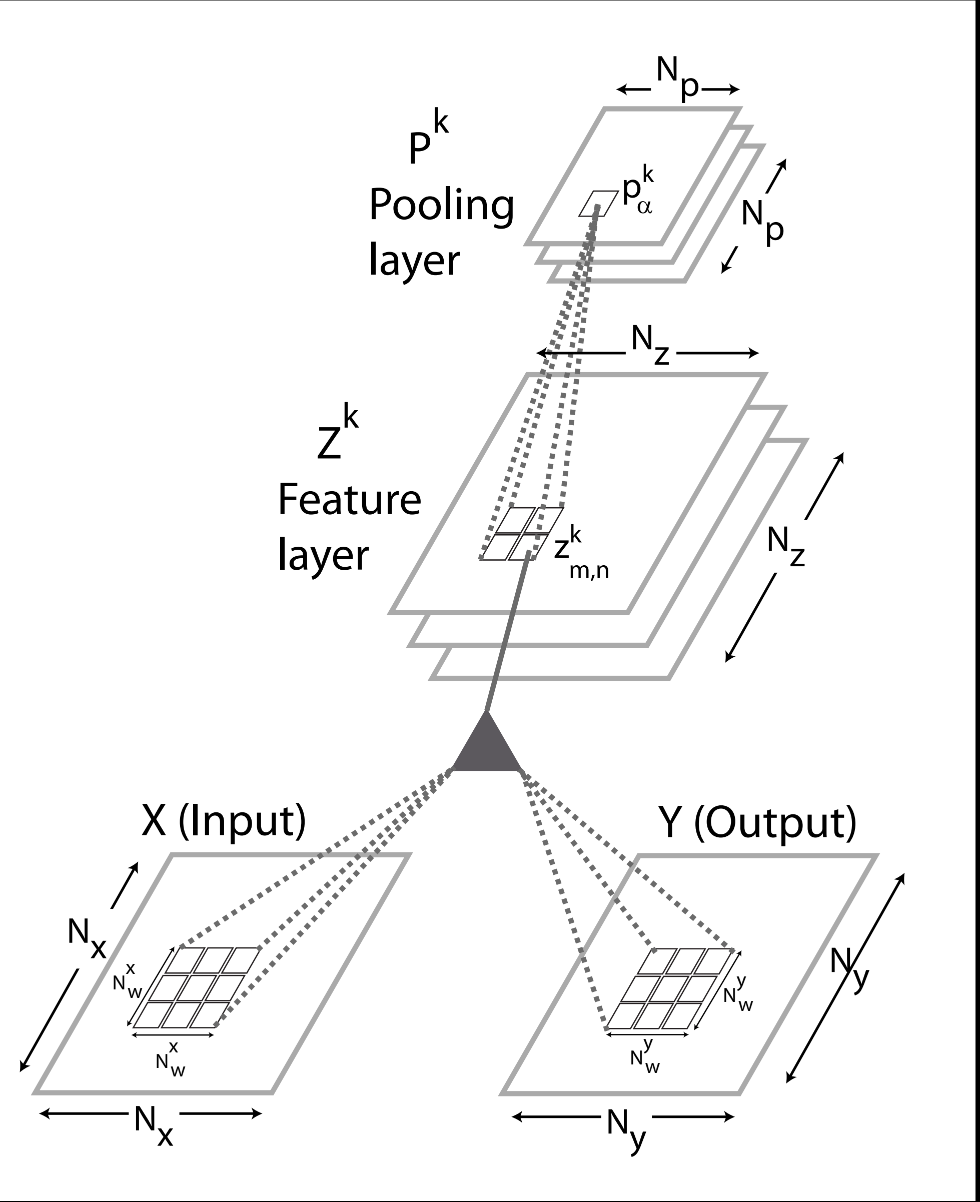
gated restricted boltzmann machines (grbm)

Memisevic & Hinton (2007)



Convolutional Gated RBM

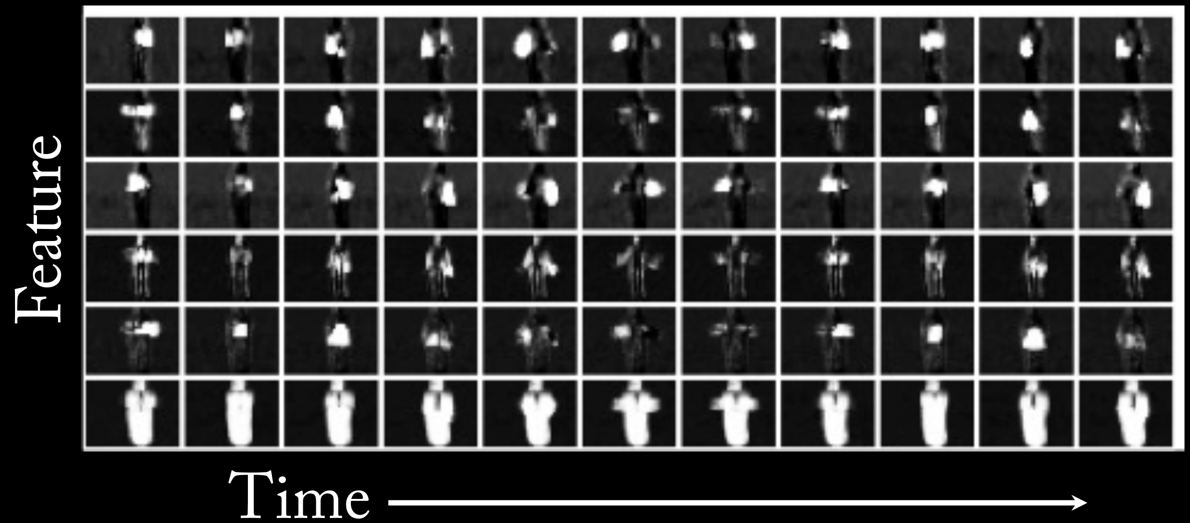
- Taylor et al. [ECCV'10]
- Local 3rd order interactions between pair of frames and features
- Inference has closed form



Visualization - Convolutional Gated RBM

- KTH actions
- Some features capture motion
- Others capture static content (e.g. edges)

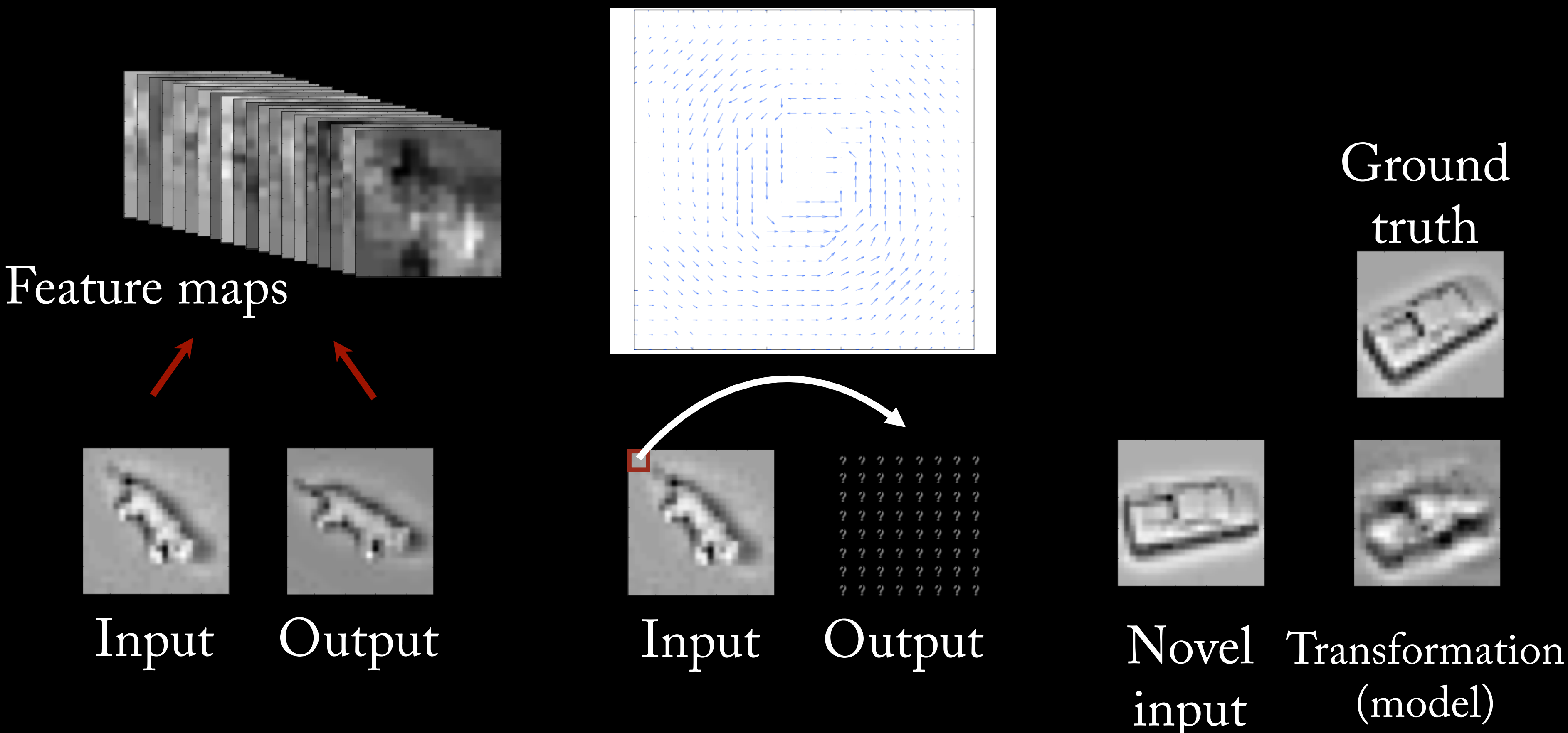
Action: Hand-clapping



(subset of features)

Taylor et al. [ECCV'10]

Visualization - Convolutional Gated RBM



[Slide: G. Taylor]