Reconstructive Sparse Code Transfer for Contour Detection and Semantic Labeling

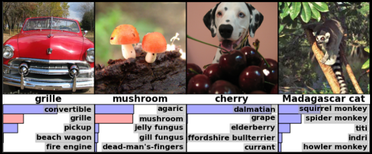
Michael Maire[1,2]   Stella X. Yu[3]   Pietro Perona[2]

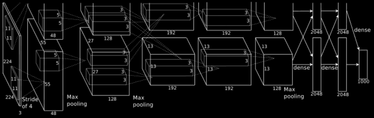
[1]TTI Chicago   [2]California Institute of Technology
[3]University of California at Berkeley / ICSI

Motivation: Deep Representations

Motivation: Deep Representations

Image Classification

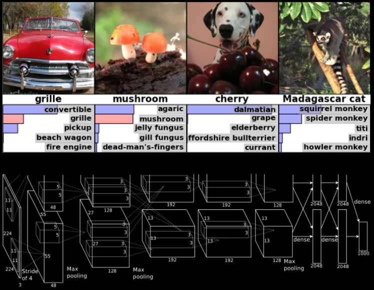




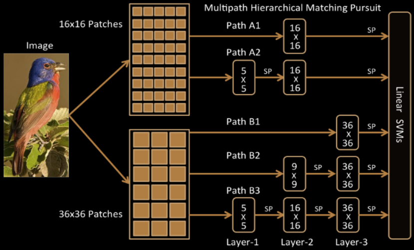
[Krizhevsky, Sutskever, and Hinton, NIPS 2012]

Motivation: Deep Representations

Image Classification



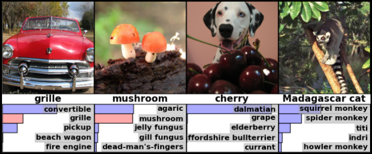
[Krizhevsky, Sutskever, and Hinton, NIPS 2012]

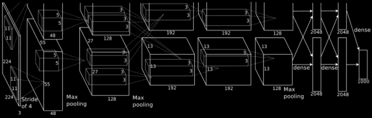


[Bo, Ren, and Fox, CVPR 2013]

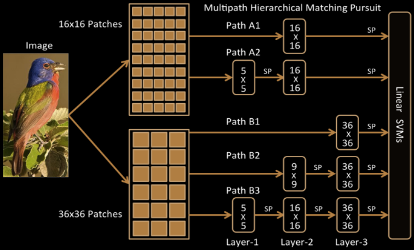
Motivation: Deep Representations

Image Classification



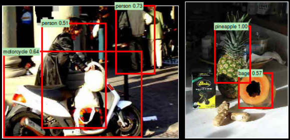


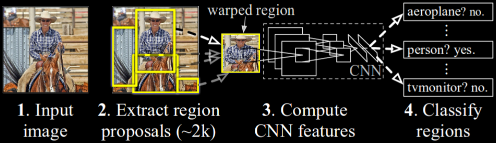
[Krizhevsky, Sutskever, and Hinton, NIPS 2012]



[Bo, Ren, and Fox, CVPR 2013]

Object Detection

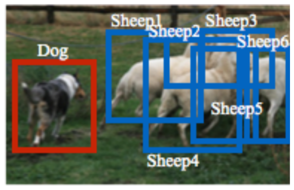




[Girshick, Donahue, Darrell, and Malik, CVPR 2014]

Deep Representations for **Semantic Labeling**

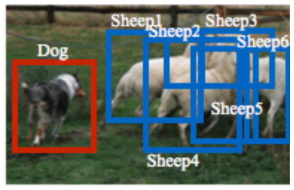
classify image



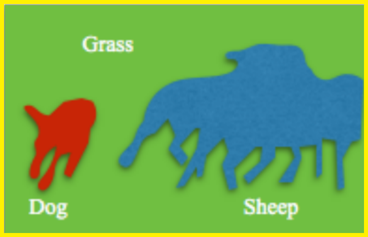
detect objects

Deep Representations for **Semantic Labeling**

classify image



detect objects

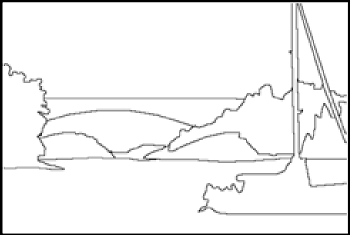


label every pixel

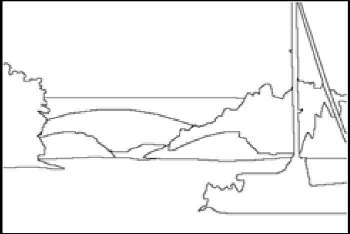
Contour Detection: Special Case

$$\rightarrow \begin{cases} 0 & \text{if in region interior} \\ 1 & \text{if on region boundary} \end{cases}$$

Contour Detection: Special Case



Contour Detection: Special Case



contour detection serves as foundation for:
segmentation, object proposals

Semantic Labeling Strategy

predict patch labels from a spatially localized multilayer slice of a deep representation

Semantic Labeling Strategy

predict patch labels from a spatially localized multilayer slice of a deep representation

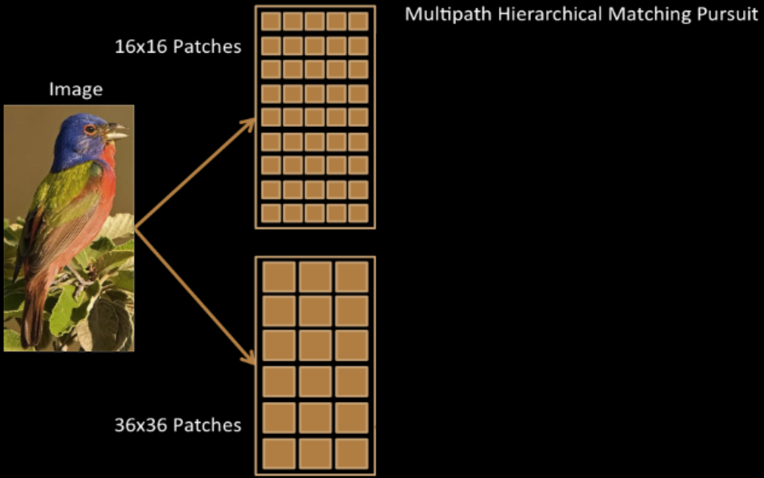
generalization of sparse reconstruction

Multipath Sparse Coding

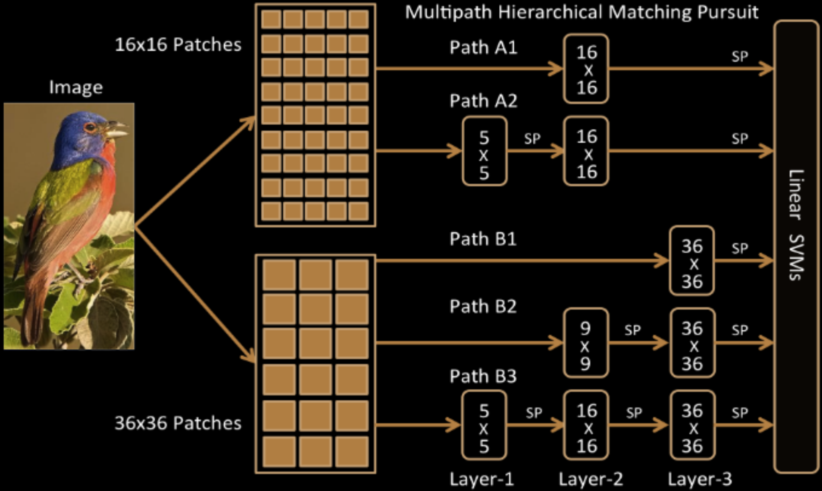
[Bo, Ren, and Fox, CVPR 2013]

Image

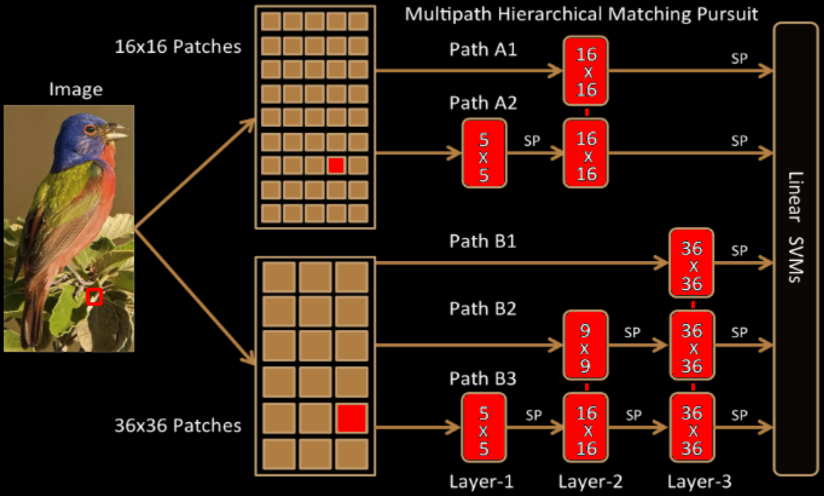
[Bo, Ren, and Fox, CVPR 2013]



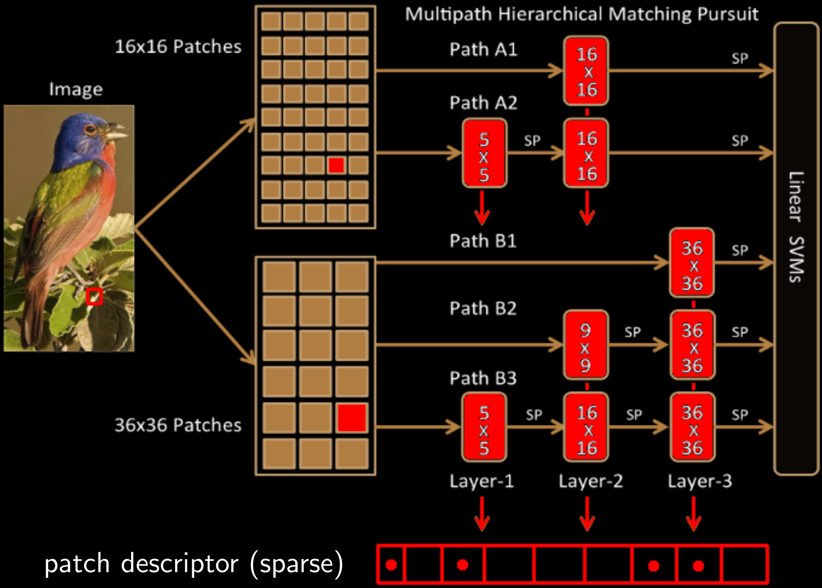
[Bo, Ren, and Fox, CVPR 2013]



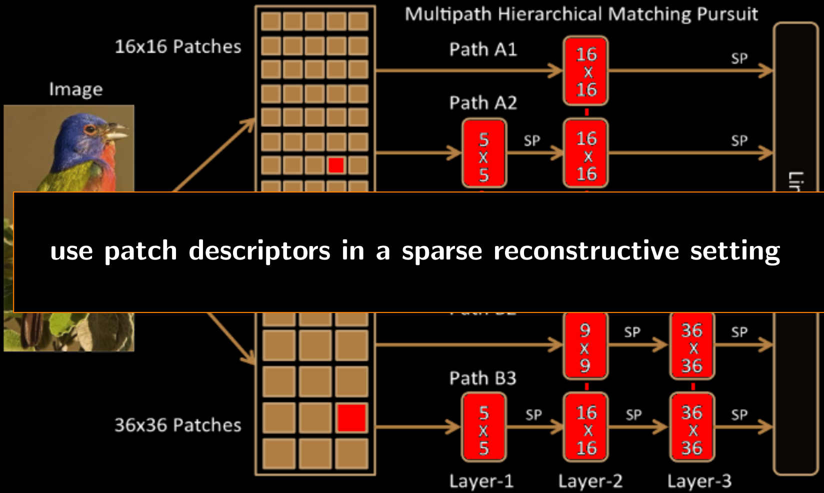
[Bo, Ren, and Fox, CVPR 2013]



Multipath Hierarchical Matching Pursuit
Image
16x16 Patches
36x36 Patches
Path A1
Path A2
Path B1
Path B2
Path B3
16 x 16
5 x 5
9 x 9
36 x 36
SP
Layer-1
Layer-2
Layer-3
Linear SVMs

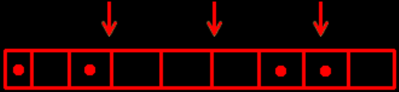


patch descriptor (sparse)



use patch descriptors in a sparse reconstructive setting

patch descriptor (sparse)



Sparse Coding & Reconstruction

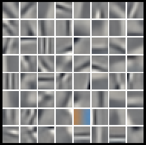
Image

Sparse Coding & Reconstruction

Image

$\Rightarrow$

Patch Dictionary



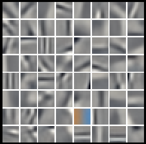
Batch OMP

Sparse Coding & Reconstruction

Image

$\Rightarrow$

Patch Dictionary



Batch OMP

 $= z_{i,0}$  $+ z_{i,1}$  $+ \ldots + z_{i,L}$  $\quad ||z_i||_0 \leq K$

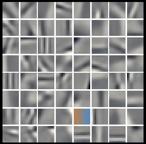
x_i $\quad d_0$ $\quad d_1$ $\quad d_L$

Sparse Coding & Reconstruction

Image

$\Rightarrow$

Patch Dictionary

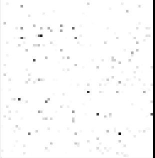


Batch OMP

$\Rightarrow$

Sparse Codes

coefficients $\rightarrow$



pixels $\rightarrow$

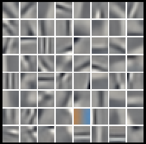
$$x_i = z_{i,0}\, d_0 + z_{i,1}\, d_1 + \dots + z_{i,L}\, d_L \qquad ||z_i||_0 \leq K$$

Sparse Coding & Reconstruction

Image

$\Rightarrow$

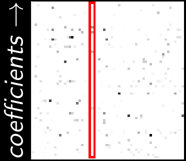
Patch Dictionary



Batch OMP

$\Rightarrow$

Sparse Codes



 $= z_{i,0}$  $+ z_{i,1}$  $+ \ldots + z_{i,L}$  $\qquad ||z_i||_0 \leq K$

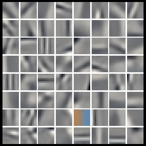
$x_i \qquad d_0 \qquad d_1 \qquad d_L$

Sparse Coding & Reconstruction

Image

$\Rightarrow$

Patch Dictionary

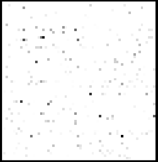


Batch OMP

$\Rightarrow$

Sparse Codes

coefficients $\rightarrow$



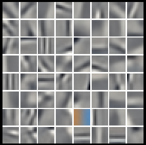
pixels $\rightarrow$

Sparse Coding & Reconstruction

Image

$\Longrightarrow$

Patch Dictionary

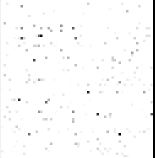


Batch OMP

$\Longrightarrow$

Sparse Codes

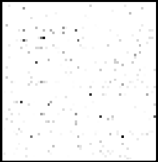
coefficients $\rightarrow$



pixels $\rightarrow$

Sparse Codes

coefficients $\rightarrow$



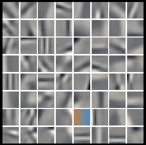
pixels $\rightarrow$

Sparse Coding & Reconstruction

Image

$\Rightarrow$

Patch Dictionary

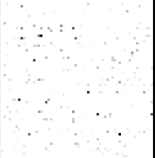


Batch OMP

$\Rightarrow$

Sparse Codes

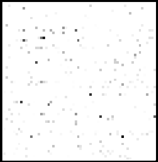
coefficients $\rightarrow$



pixels $\rightarrow$

Sparse Codes

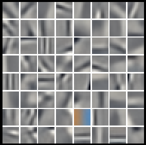
coefficients $\rightarrow$



pixels $\rightarrow$

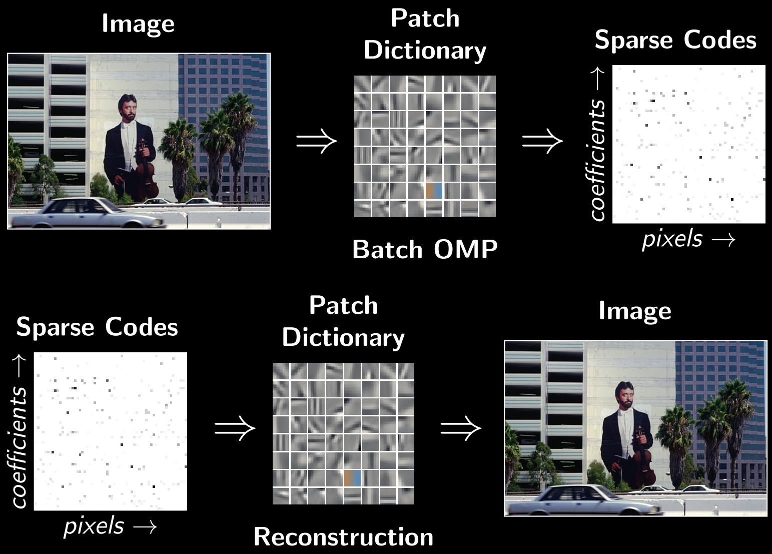
$\Rightarrow$

Patch Dictionary



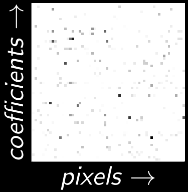
Reconstruction

Sparse Coding & Reconstruction



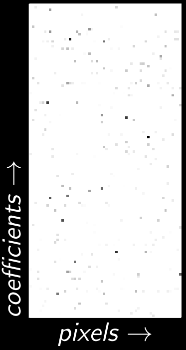
Reconstructive Sparse Code Transfer

Sparse Codes



Reconstructive Sparse Code Transfer

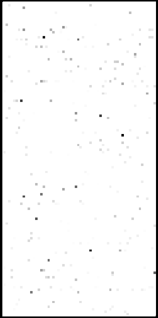
Rectified Sparse Codes



$$z_i \leftarrow \left[\max(z_i^T, 0), \max(-z_i^T, 0), 1\right]^T$$

Reconstructive Sparse Code Transfer

Rectified Sparse Codes

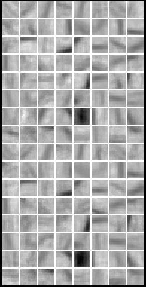


coefficients $\rightarrow$

pixels $\rightarrow$

$\Rightarrow$

Transfer Dictionary

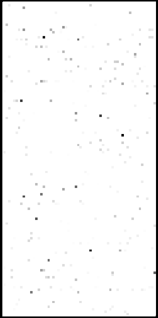


Reconstruction

$$z_i \leftarrow \left[\max(z_i^T, 0), \max(-z_i^T, 0), 1\right]^T$$

Reconstructive Sparse Code Transfer

Rectified Sparse Codes

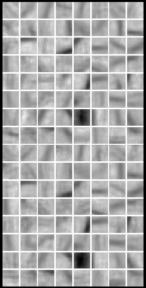


coefficients $\rightarrow$

pixels $\rightarrow$

$\Rightarrow$

Transfer Dictionary



Reconstruction

$\Rightarrow$

Contour Detection

$$z_i \leftarrow \left[\max(z_i^T, 0), \max(-z_i^T, 0), 1\right]^T$$

Sparse Coding: Generative Training

Sparse Coding: Generative Training

- sample patches $X = [x_0, x_1, ...]$ from training images

Sparse Coding: Generative Training

- sample patches $X = [x_0, x_1, ...]$ from training images
- learn:
  - sparse representations $Z = [z_0, z_1, ...]$
  - dictionary $D = [d_0, d_1, \ldots, d_{L-1}]$

Sparse Coding: Generative Training

- sample patches $X = [x_0, x_1, ...]$ from training images
- learn:
  - sparse representations $Z = [z_0, z_1, ...]$
  - dictionary $D = [d_0, d_1, \ldots, d_{L-1}]$
- MI-KSVD finds an approximate solution to:

$$\underset{D,\ Z}{\operatorname{argmin}} \left[||X - DZ||_F^2 + \lambda \sum_{i=0}^{L-1} \sum_{j=0, j \neq i}^{L-1} |d_i^T d_j| \right]$$

$$s.t.\ \forall i,\ ||d_i||_2 = 1 \text{ and } \forall n,\ ||z_n||_0 \leq K$$

Sparse Coding: Generative Training

- sample patches $X = [x_0, x_1, ...]$ from training images
- learn:
  - sparse representations $Z = [z_0, z_1, ...]$
  - dictionary $D = [d_0, d_1, \ldots, d_{L-1}]$
- MI-KSVD finds an approximate solution to:

$$\underset{D,\ Z}{\operatorname{argmin}} \left[||X - DZ||_F^2 + \lambda \sum_{i=0}^{L-1} \sum_{j=0, j \neq i}^{L-1} |d_i^T d_j| \right]$$

$$s.t.\ \forall i,\ ||d_i||_2 = 1 \text{ and } \forall n,\ ||z_n||_0 \leq K$$

- encode patch $x \in \Re^{m \cdot m \cdot c}$ as $z \in \Re^L$:

$$\underset{z}{\operatorname{argmin}} ||x - Dz||^2 \quad s.t.\ ||z||_0 \leq K$$

Transfer Dictionary: Discriminative Training

Transfer Dictionary: Discriminative Training

- D maps sparse vector $z \in \Re^L$ to patch $P \in \Re^{m \cdot m \cdot c}$

Transfer Dictionary: Discriminative Training

- D maps sparse vector $z \in \Re^L$ to patch $P \in \Re^{m \cdot m \cdot c}$
- replace D with function $F(\cdot)$

Transfer Dictionary: Discriminative Training

- D maps sparse vector $z \in \Re^L$ to patch $P \in \Re^{m \cdot m \cdot c}$
- replace D with function $F(\cdot)$
- $F(z)$ predicts groundtruth labeling $G \in \Re^{m \cdot m \cdot h}$

Transfer Dictionary: Discriminative Training

- D maps sparse vector $z \in \Re^{L}$ to patch $P \in \Re^{m \cdot m \cdot c}$
- replace D with function $F(\cdot)$
- $F(z)$ predicts groundtruth labeling $G \in \Re^{m \cdot m \cdot h}$
- patch-level discriminative training:

Transfer Dictionary: Discriminative Training

- D maps sparse vector $z \in \Re^L$ to patch $P \in \Re^{m \cdot m \cdot c}$
- replace D with function $F(\cdot)$
- $F(z)$ predicts groundtruth labeling $G \in \Re^{m \cdot m \cdot h}$
- patch-level discriminative training:
  - sample codes $\{z_0, z_1, \ldots\}$ and corresponding groundtruth patches $\{g_0, g_1, \ldots\}$

Transfer Dictionary: Discriminative Training

- D maps sparse vector $z \in \Re^L$ to patch $P \in \Re^{m \cdot m \cdot c}$
- replace D with function $F(\cdot)$
- $F(z)$ predicts groundtruth labeling $G \in \Re^{m \cdot m \cdot h}$
- patch-level discriminative training:
  - sample codes $\{z_0, z_1, \ldots\}$ and corresponding groundtruth patches $\{g_0, g_1, \ldots\}$
  - split $F(\cdot)$ into independently trained predictors $[f_0, f_1, ..., f_{(m^2 h-1)}]$, one for each output element

Transfer Dictionary: Discriminative Training

- D maps sparse vector $z \in \Re^L$ to patch $P \in \Re^{m \cdot m \cdot c}$
- replace D with function $F(\cdot)$
- $F(z)$ predicts groundtruth labeling $G \in \Re^{m \cdot m \cdot h}$
- patch-level discriminative training:
    - sample codes $\{z_0, z_1, \ldots\}$ and corresponding groundtruth patches $\{g_0, g_1, \ldots\}$
    - split $F(\cdot)$ into independently trained predictors $[f_0, f_1, ..., f_{(m^2 h-1)}]$, one for each output element
    - we want:

$$f_j(z_i) \approx g_i[j] \quad \forall i, j$$

Transfer Dictionary: Discriminative Training

- D maps sparse vector $z \in \Re^L$ to patch $P \in \Re^{m \cdot m \cdot c}$
- replace D with function $F(\cdot)$
- $F(z)$ predicts groundtruth labeling $G \in \Re^{m \cdot m \cdot h}$
- patch-level discriminative training:
  - sample codes $\{z_0, z_1, \ldots\}$ and corresponding groundtruth patches $\{g_0, g_1, \ldots\}$
  - split $F(\cdot)$ into independently trained predictors $[f_0, f_1, ..., f_{(m^2 h-1)}]$, one for each output element
  - we want:

$$f_j(z_i) \approx g_i[j] \quad \forall i, j$$

  - train each $f_j(\cdot)$ using logistic regression

Transfer Dictionary: Discriminative Training

- D maps sparse vector $z \in \Re^L$ to patch $P \in \Re^{m \cdot m \cdot c}$
- replace D with function $F(\cdot)$
- $F(z)$ predicts groundtruth labeling $G \in \Re^{m \cdot m \cdot h}$
- patch-level discriminative training:
  - sample codes $\{z_0, z_1, \ldots\}$ and corresponding groundtruth patches $\{g_0, g_1, \ldots\}$
  - split $F(\cdot)$ into independently trained predictors $[f_0, f_1, ..., f_{(m^2 h-1)}]$, one for each output element
  - we want:
  $$f_j(z_i) \approx g_i[j] \quad \forall i, j$$
  - train each $f_j(\cdot)$ using logistic regression
- replace z with concatenated multipath codes

Multiple Scales

Multiple Scales

×

Layer 1 Dictionaries

5x5, 64 atoms

11x11, 64 atoms

5x5, 512 atoms

11x11, 512 atoms

21x21, 512 atoms

31x31, 512 atoms

Multiple Scales

×

Layer 1 Dictionaries

5x5, 64 atoms

11x11, 64 atoms

5x5, 512 atoms

11x11, 512 atoms

21x21, 512 atoms

31x31, 512 atoms

pooling →

Layer 2 Dictionaries

5x5, 512 atoms

5x5, 512 atoms

Multiple Scales

×

Layer 1 Dictionaries

5x5, 64 atoms
11x11, 64 atoms

5x5, 512 atoms
11x11, 512 atoms
21x21, 512 atoms
31x31, 512 atoms

$\longrightarrow$ *pooling* $\longrightarrow$

Layer 2 Dictionaries

5x5, 512 atoms
5x5, 512 atoms

$\Downarrow$

$\Longrightarrow$ *rectify, upsample, concatenate sparse activation maps*

$\swarrow$

Multiple Scales

×

Layer 1 Dictionaries

- 5x5, 64 atoms
- 11x11, 64 atoms

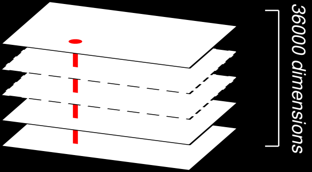
- 5x5, 512 atoms
- 11x11, 512 atoms
- 21x21, 512 atoms
- 31x31, 512 atoms

pooling

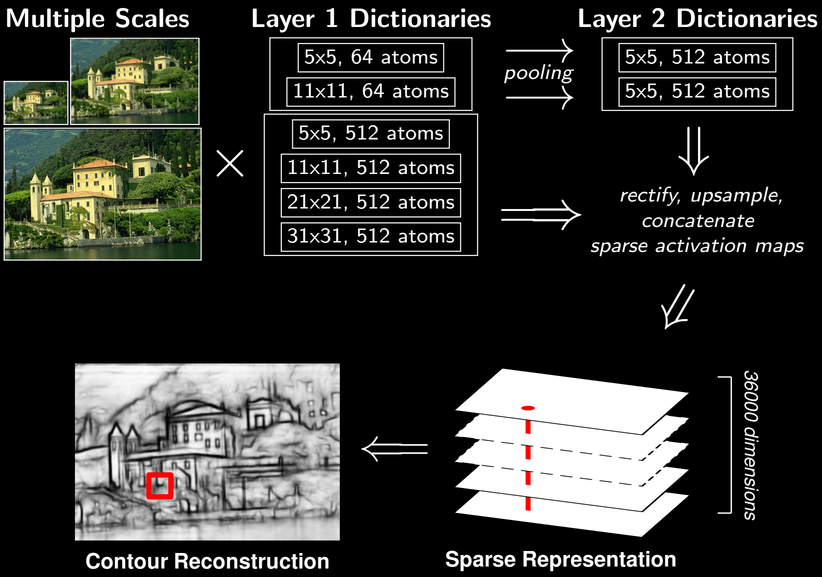
Layer 2 Dictionaries

- 5x5, 512 atoms
- 5x5, 512 atoms

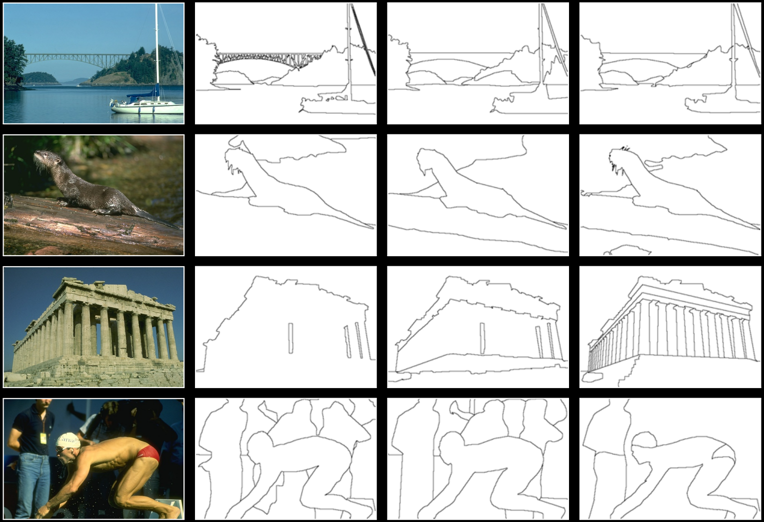
rectify, upsample,
concatenate
sparse activation maps



Sparse Representation


Multiple Scales
Layer 1 Dictionaries
5x5, 64 atoms
11x11, 64 atoms
5x5, 512 atoms
11x11, 512 atoms
21x21, 512 atoms
31x31, 512 atoms
pooling
Layer 2 Dictionaries
5x5, 512 atoms
5x5, 512 atoms
rectify, upsample,
concatenate
sparse activation maps
36000 dimensions
Sparse Representation
Contour Reconstruction

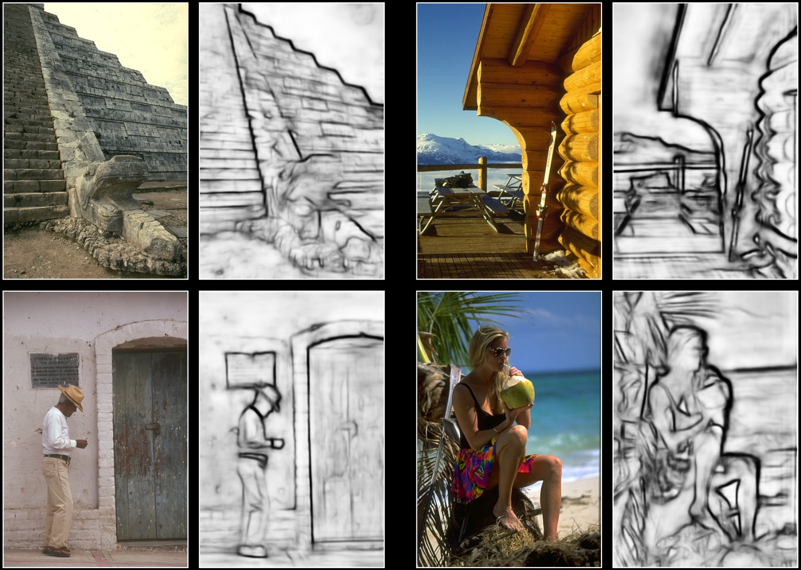
Contour Detection Groundtruth



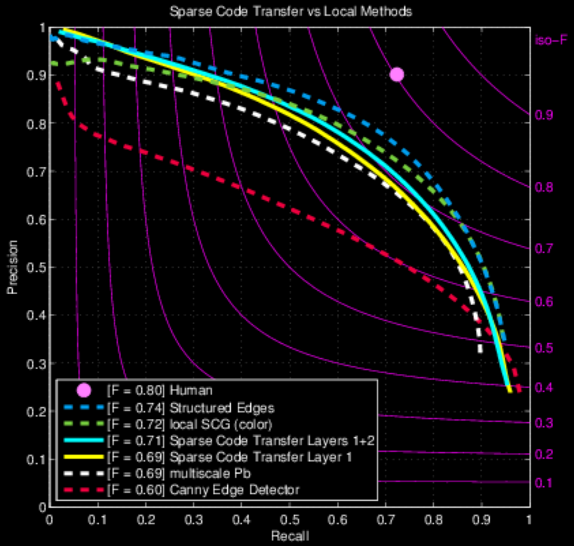
[Martin, Fowlkes, Tal, and Malik, ICCV 2001]

Contour Detection Results

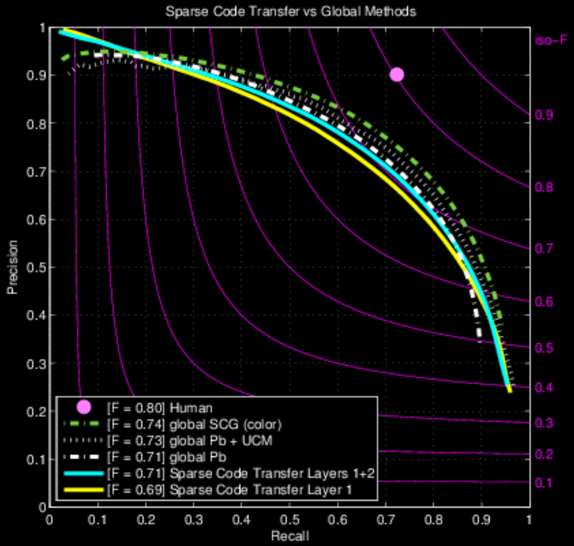
Contour Detection Results



Contour Detection Performance



Contour Detection Performance



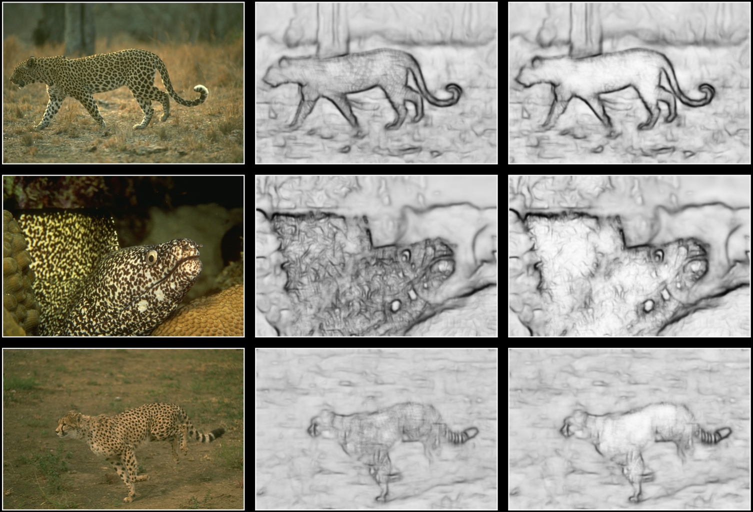
Contour Detection Performance

	Performance Metric			Hand-Designed		Spectral
	ODS F	OIS F	AP	Features?	Filters?	Globalization?
Human	0.80	0.80	—	—	—	—
Structured Edges	0.74	0.76	0.78	yes	no	no
local SCG (color)	0.72	0.74	0.75	no	yes	no
Sparse Code Transfer Layers 1+2	**0.71**	**0.72**	**0.74**	**no**	**no**	**no**
Sparse Code Transfer Layer 1	**0.69**	**0.71**	**0.72**	**no**	**no**	**no**
local SCG (gray)	0.69	0.71	0.71	no	yes	no
multiscale Pb	0.69	0.71	0.68	yes	yes	no
Canny Edge Detector	0.60	0.63	0.58	yes	yes	no
global SCG (color)	0.74	0.76	0.77	yes	yes	yes
global Pb + UCM	0.73	0.76	0.73	yes	yes	yes + UCM
global Pb	0.71	0.74	0.65	yes	yes	yes

Sparse Code Transfer:

- Performance competitive with top approaches
- Both representation and classifier are learned
- Free from reliance on hand-designed features or filters

Texture and Network Depth

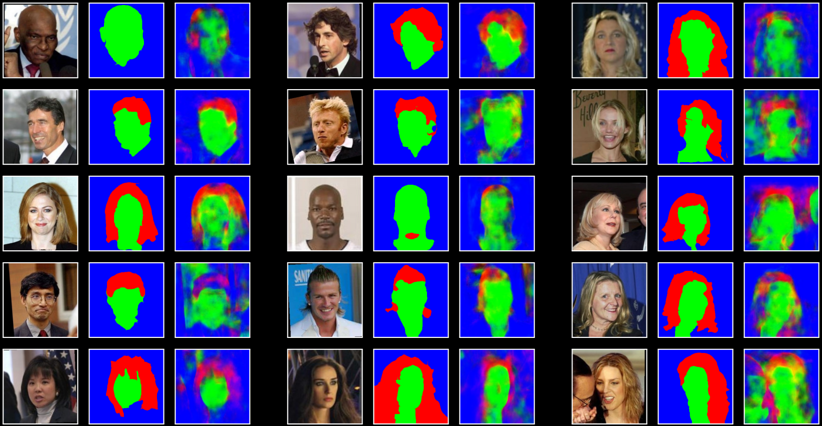


Layer 1 Layers 1+2

Texture and Network Depth

Layer 1 | Layers 1+2

Semantic Labeling



hair skin background

Labeled Faces in the Wild Dataset [Kae, Sohn, Lee, Learned-Miller, CVPR 2013]

Summary

Summary

- Semantic labeling using spatially localized slices of deep representations

Summary

- Semantic labeling using spatially localized slices of deep representations
- Sparse reconstruction generalized to multipath networks

Summary

- Semantic labeling using spatially localized slices of deep representations
- Sparse reconstruction generalized to multipath networks
- Transfer learning perspective:
  - Generatively trained patch representation
  - Task-specific discriminately trained transfer classifier

Summary

- Semantic labeling using spatially localized slices of deep representations
- Sparse reconstruction generalized to multipath networks
- Transfer learning perspective:
    - Generatively trained patch representation
    - Task-specific discriminately trained transfer classifier
- Pure learning for high performance contour detection

Summary

- Semantic labeling using spatially localized slices of deep representations
- Sparse reconstruction generalized to multipath networks
- Transfer learning perspective:
  - Generatively trained patch representation
  - Task-specific discriminately trained transfer classifier
- Pure learning for high performance contour detection
- Contours obtained as a byproduct of deep representations

Summary

- Semantic labeling using spatially localized slices of deep representations
- Sparse reconstruction generalized to multipath networks
- Transfer learning perspective:
  - Generatively trained patch representation
  - Task-specific discriminately trained transfer classifier
- Pure learning for high performance contour detection
- Contours obtained as a byproduct of deep representations

Thank You!