Domain Transfer Learning for Vision Applications:
Transfer via Shared Representations

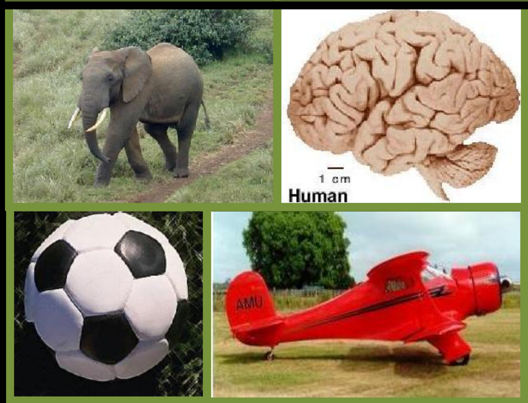
Kate Saenko

Harvard University

University of California, Berkeley

Do we need domain transfer? Isn't the object recognition field making great progress?

Caltech 101 images



PASCAL images

% accuracy on Caltech101

% avg. precision on Pascal

What's been driving the success behind object recognition

Invariant descriptors

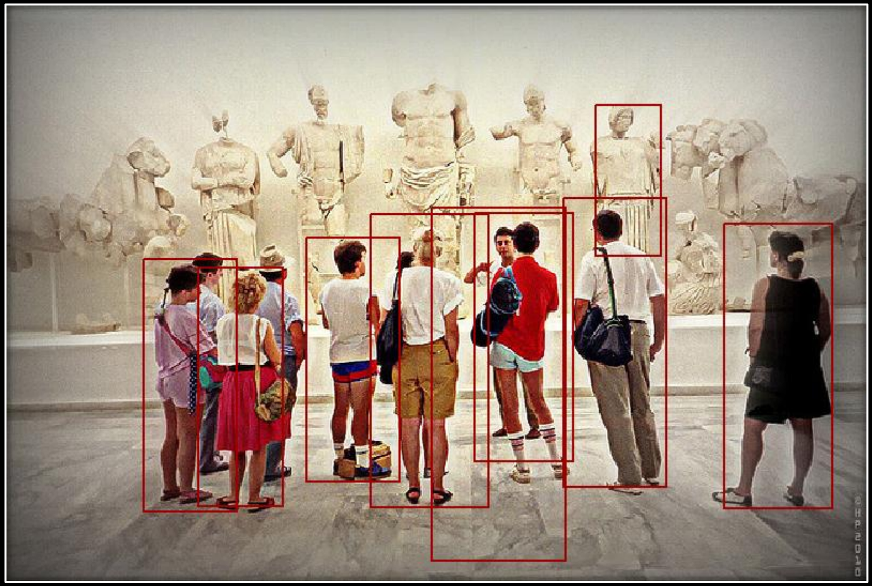
+

Machine learning

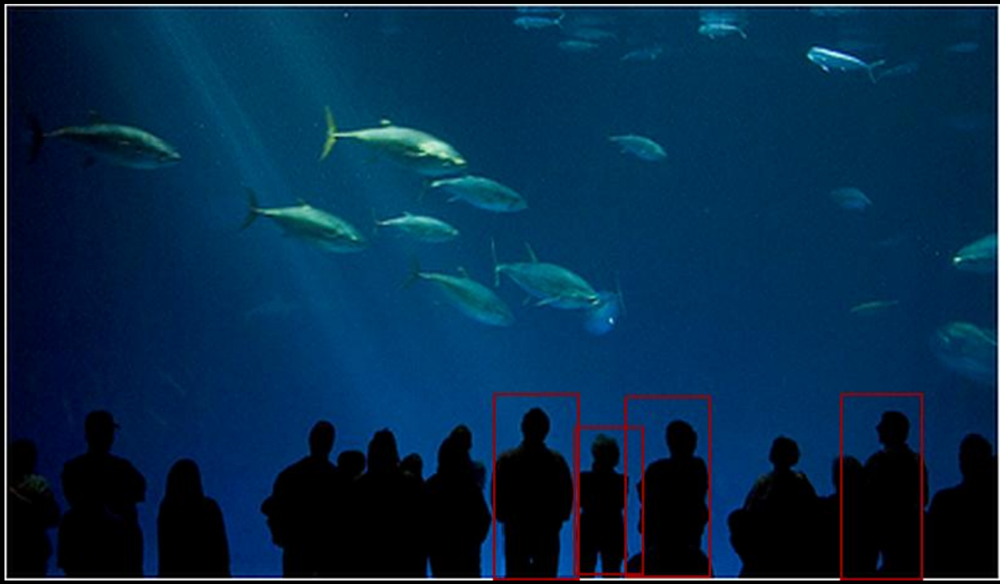
+

Big data

Problem:
PASCAL Person detector fails in a new domain



Flick image, similar to PASCAL
Good



New domain, “aquarium”
Poor

What to do? Train on more data?

Invariant descriptors

+

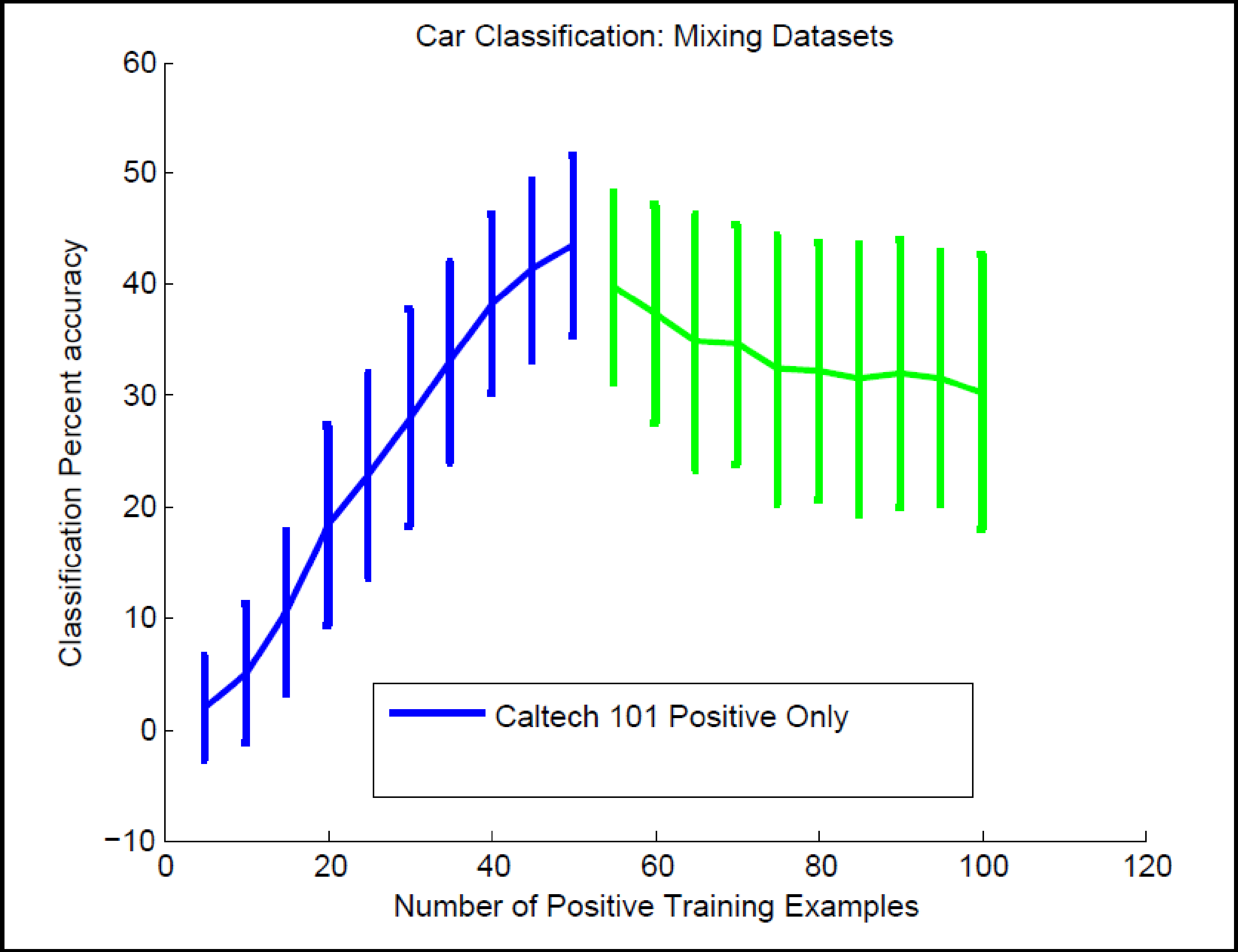
Machine learning

+

Big data

Dilemma: just adding more data does not help

“More data ain’t always better!”



Toralba and Efros, Unbiased look at dataset bias. CVPR’11

What to do? Hand tune features to each new domain?

Invariant descriptors

+

Machine learning

+

Big data

Dilemma: Hard to predict what will change in new domain

high quality

low quality

daylight

sunset

posed

“in the wild”

art

surveillance

Solution: Learn what's changed!

Invariant descriptors

+

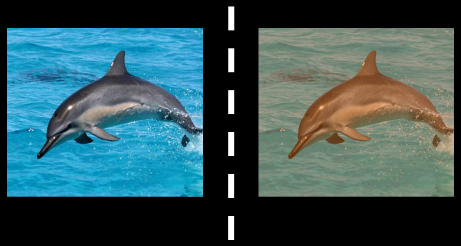
Machine learning

+

Big data

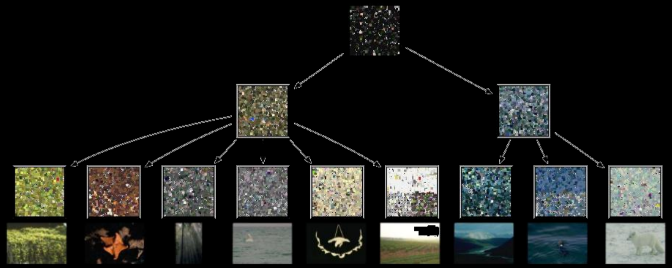
Outline

When statistics of the data change: Cross-domain methods that adapt features



- domain transform [Saenko'10]
- asymmetric transform [Kulis'11]
- manifold walks [Gopalan'11]

When labels are expensive: Cross-knowledge methods that share features



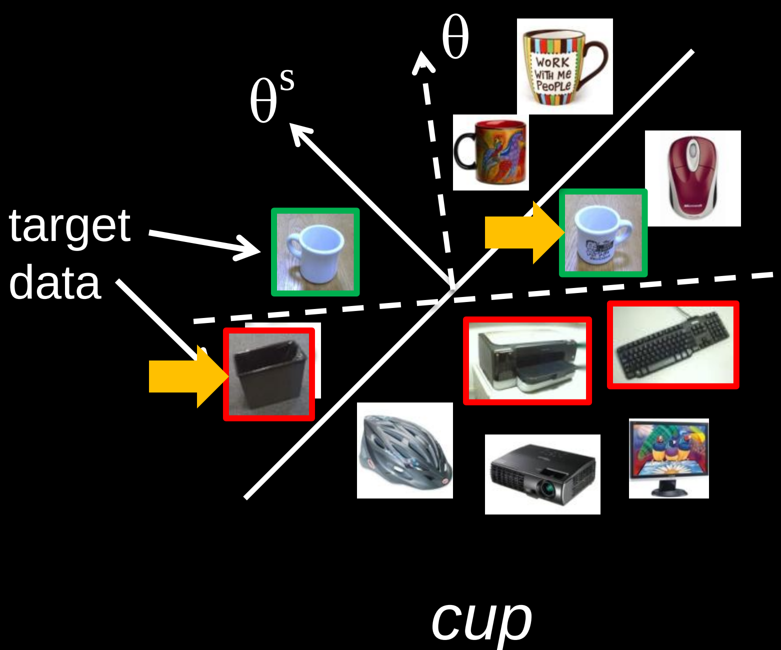
- sharing features across tasks [Torralba'04], [Quattoni'08]
- visual taxonomies [Bart'08]

How can we learn what's changed between two domains?

- Use target data (labeled or unlabeled) to adapt
- Methods vary in what they adapt:
  - Some adapt parameters: learn new classifier parameters, keep same features
  - Some adapt features: learn new feature space, keep same classifier

Adapt parameters of classifier

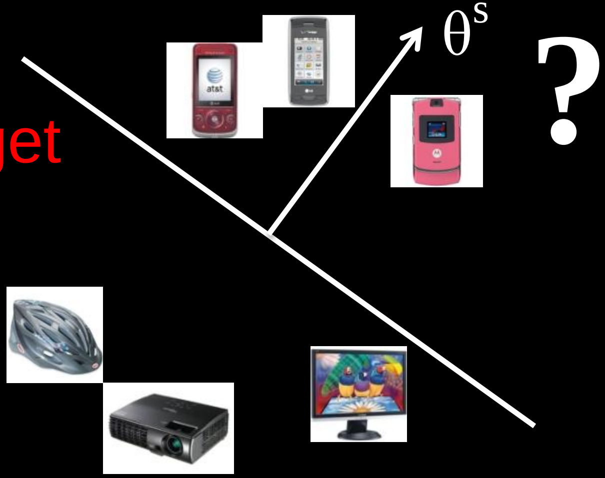
- Example: adapt SVM parameter θ^s for *cup*



- combination of source- and domain-specific classifier parameters, e.g. [Daumé ’07]
- adapt SVM parameters , e.g. [Yang’07] [Duan’09] [Duan’10]

Problem 1:
what about new categories?

- Example: adapt SVM parameter θ^s for *phone*



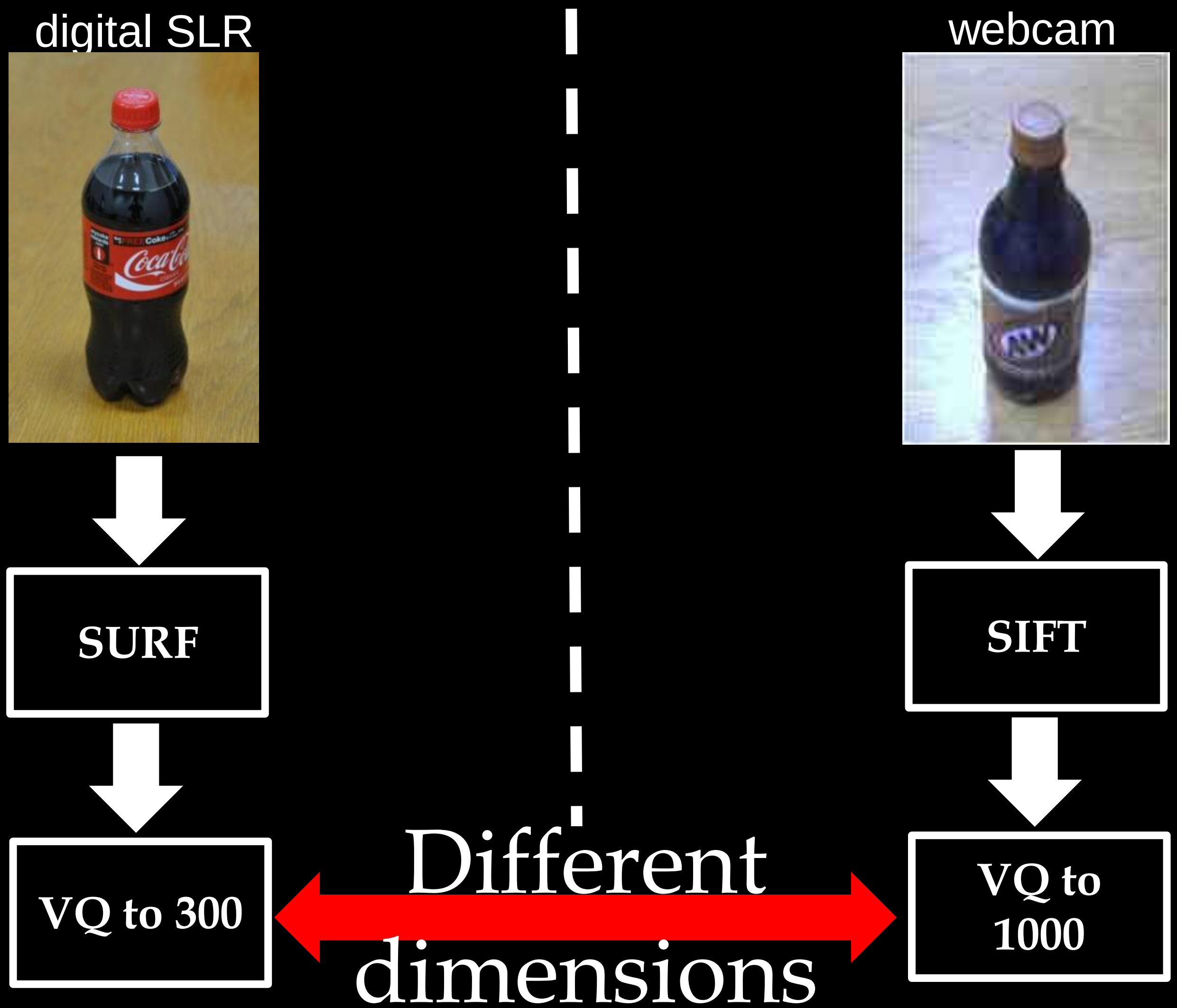
no target data!

phone

- supervised parameter adaptation learns θ per class
- what if we do not have labels for all categories?

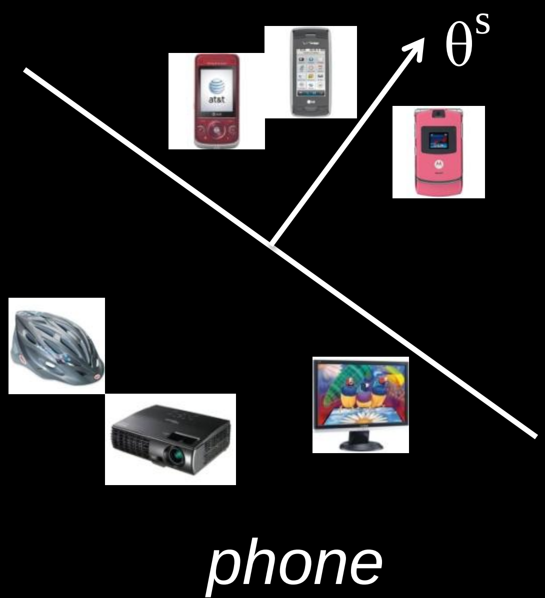
Problem 2: heterogeneous domains

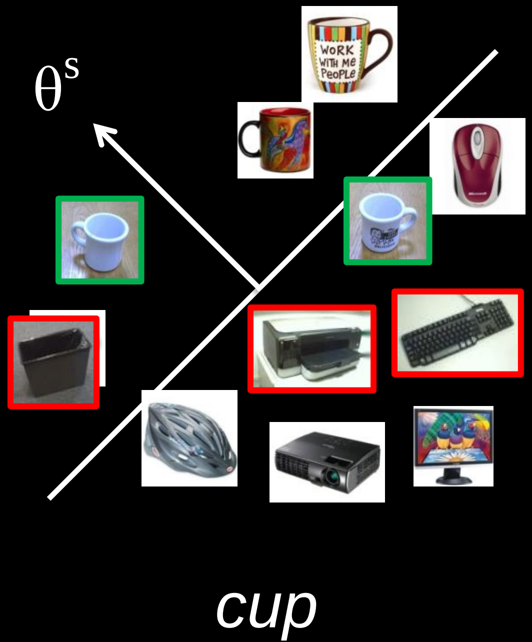
Solution: work in feature space!



Keep classifier, adapt features

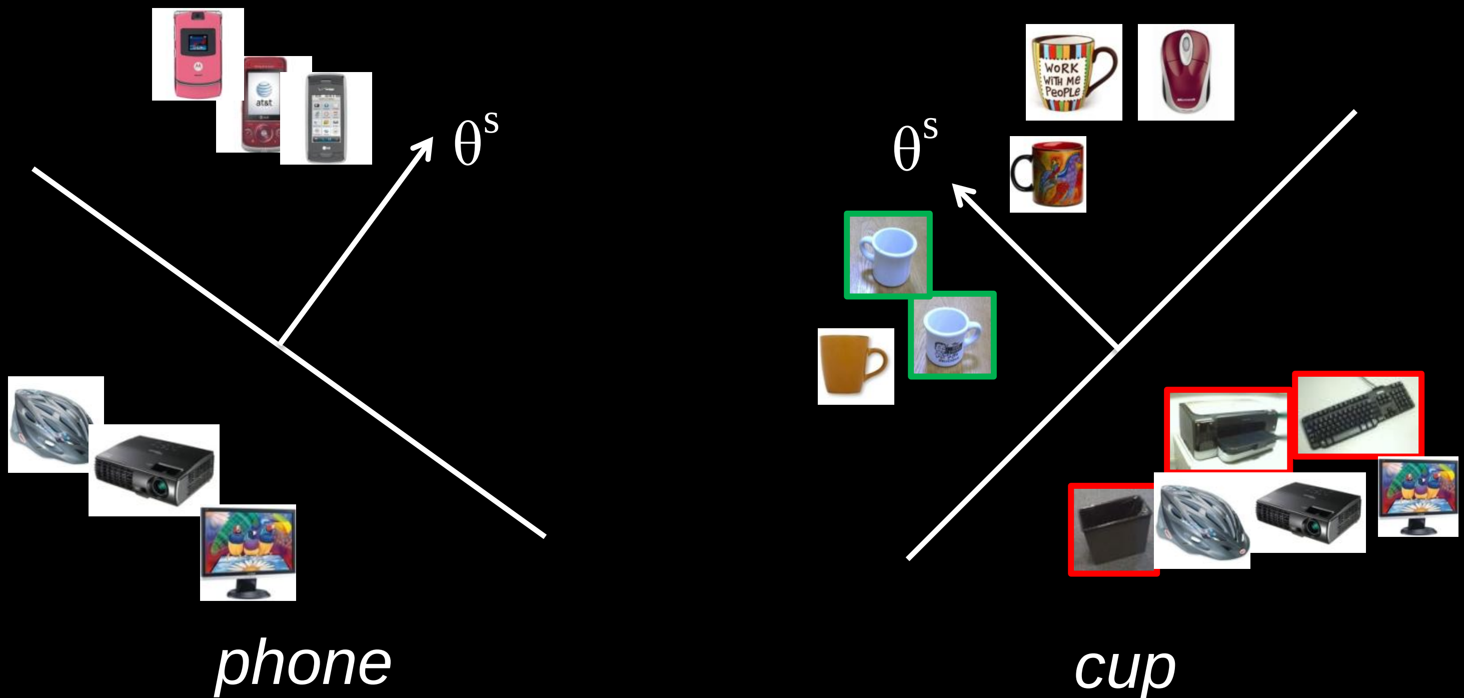
- Old feature space:





Keep classifier, adapt features

- New feature space:



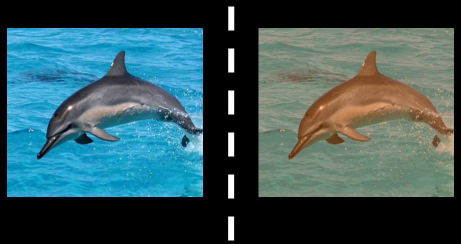
Keep classifier, adapt features

- re-weight features, or transform features
  - match test data distribution, e.g. [Farhadi'08] ,[Dai'08]
  - satisfy similarity constraints [Saenko '10], [Kulis'11]
  - exploit manifold structure [Gopalan'11]

Outline

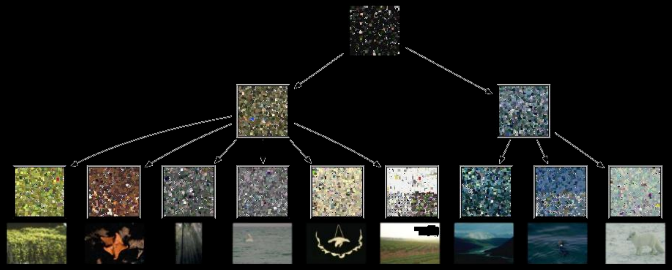
When statistics of the data change: Cross-domain methods that adapt features

- domain transform [Saenko'10]
- asymmetric transform [Kulis'11]
- manifold walks [Gopalan'11]



When labels are expensive: Cross-knowledge (same-domain) methods

- sharing features across tasks [Torralba'04], [Quattoni'08]
- visual taxonomies [Bart'08]

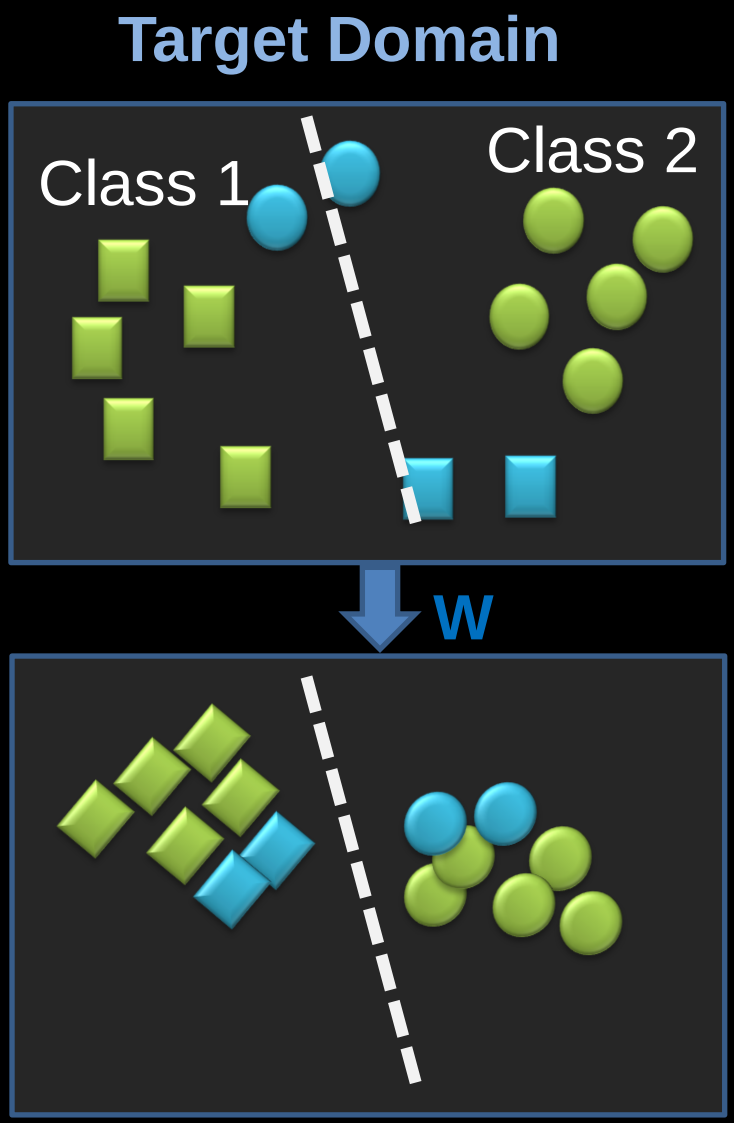


[Saenko'10] Transform-based Domain Adaptation

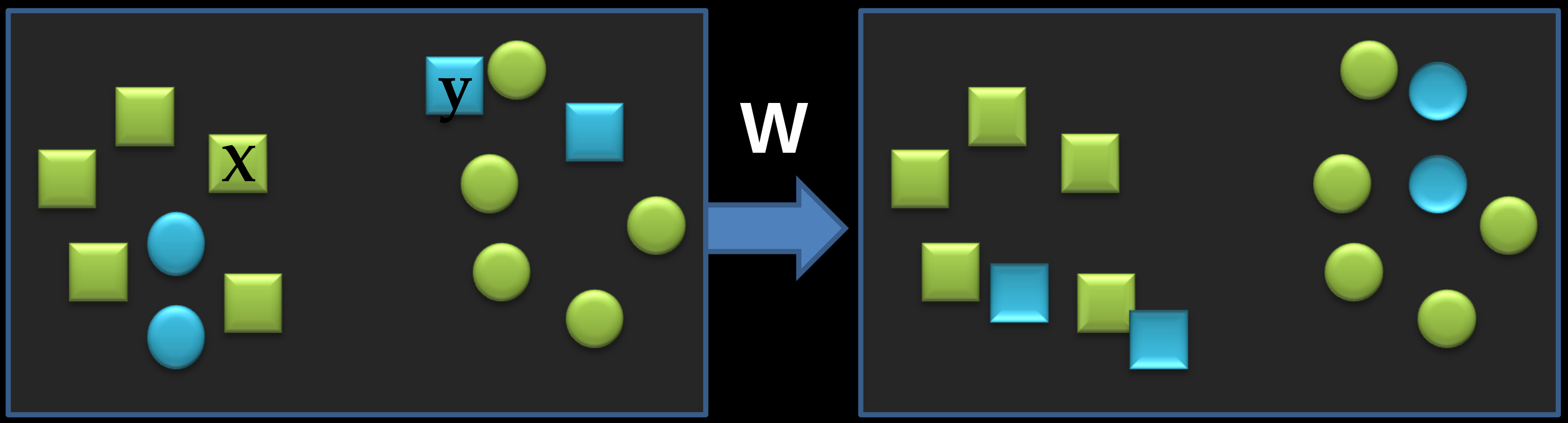
Previous methods' drawbacks

- ✘ cannot apply learned shift to new categories
- ✘ cannot handle new features

We can do both by learning *transformations*



[Saenko'10] Model Details



- Learn a *linear* transformation to map points from one domain to another
  - Call this transformation W
  - Matrices of source and target points:

$$X = [x_1, \ldots, x_{n_A}] \qquad Y = [y_1, \ldots, y_{n_B}]$$

[Saenko'10] Regularized Objective Function

- Minimize a linear combination of sum of loss functions and a regularizer:

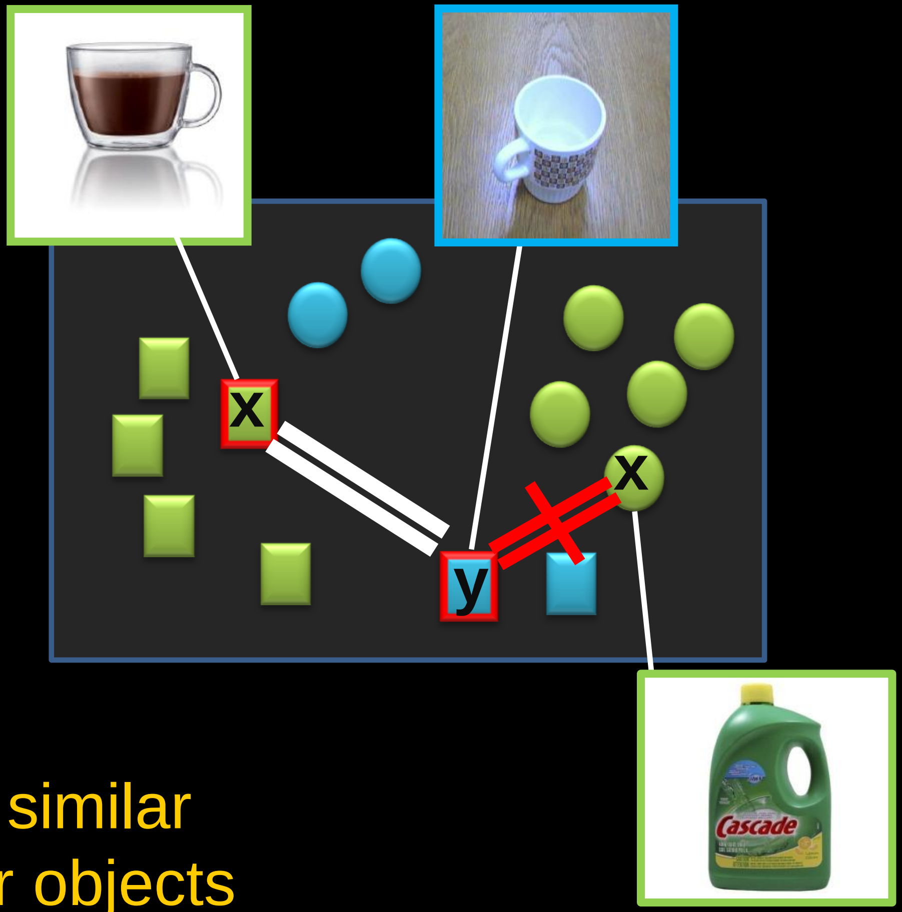
$$\min_W r(W) + \lambda \sum_{i=1}^{m} c_i(X^T W Y)$$

- Choice of regularizer affects form of W
  - LogDet norm leads to symmetric p.s.d W
  - Other choices possible

[Saenko'10]
Loss Functions

Consider distance between a point *x* from the source and *y* from the target:

$$(x-y)^T \underbrace{W}_{\text{transformation}} (x-y)$$



Force distance to be "small" for similar objects and "large" for dissimilar objects

$$Loss = \begin{cases} (\max(0, (x-y)^T W (x-y)) - u)^2 & \text{if } x \text{ and } y \text{ similar,} \\ (\max(0, l - (x-y)^T W (x-y)))^2 & \text{if } x \text{ and } y \text{ dissimilar.} \end{cases}$$

[Saenko'10] Loss Functions

- Input to problem includes a collection of *m* loss functions

$$c_1, \ldots, c_m$$

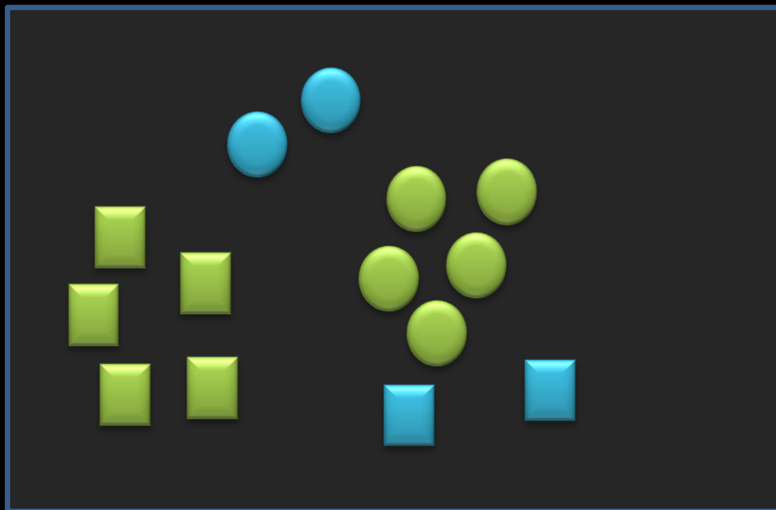
- General assumption: loss functions depend on data only through inner product matrix

$$X^T W Y$$

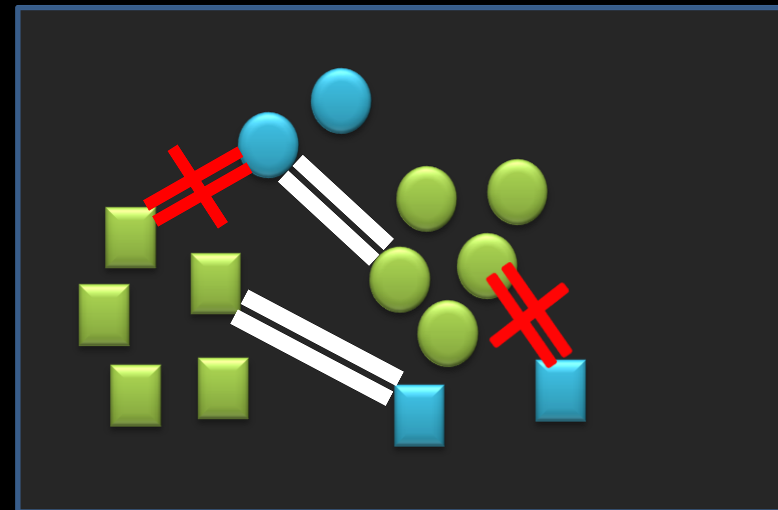
$$c_1(X^T W Y), c_2(X^T W Y), \ldots, c_m(X^T W Y)$$

[Saenko'10] Summary of approach

Train Time

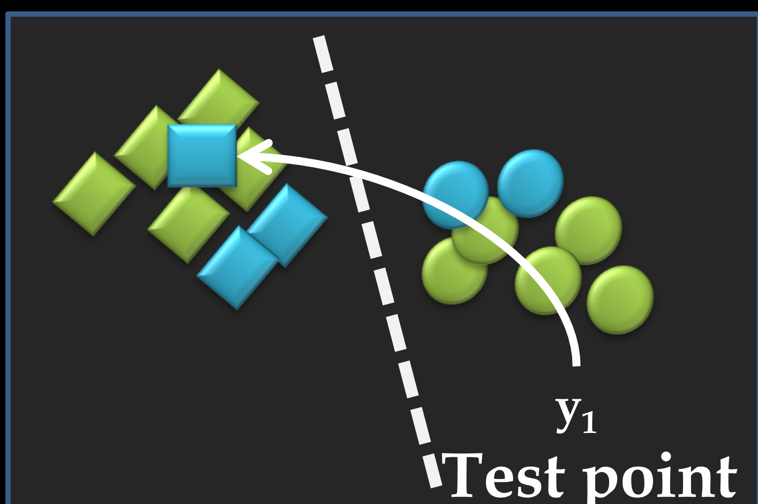


1. Multi-Domain Data

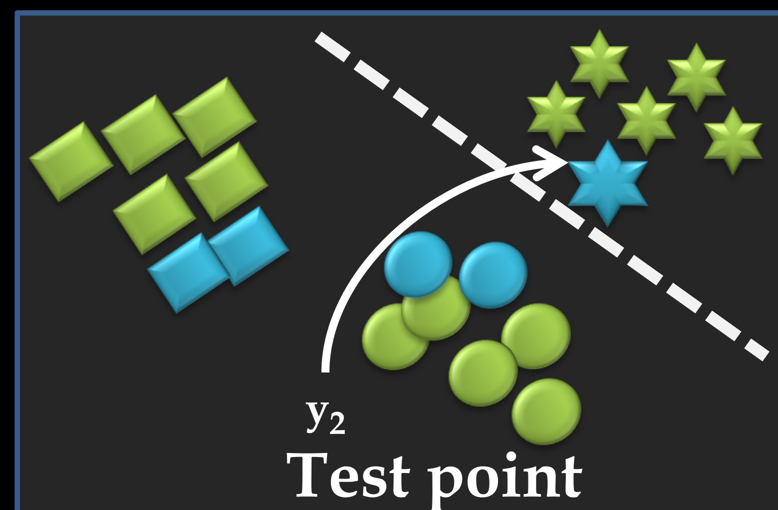


2. Generate Constraints, Learn W

Test Time



3. Map via W, classify



4. Apply W to New Categories

Limitations of symmetric transforms

[Saenko'10] essentially used metric learning:

- enforces symmetric transforms

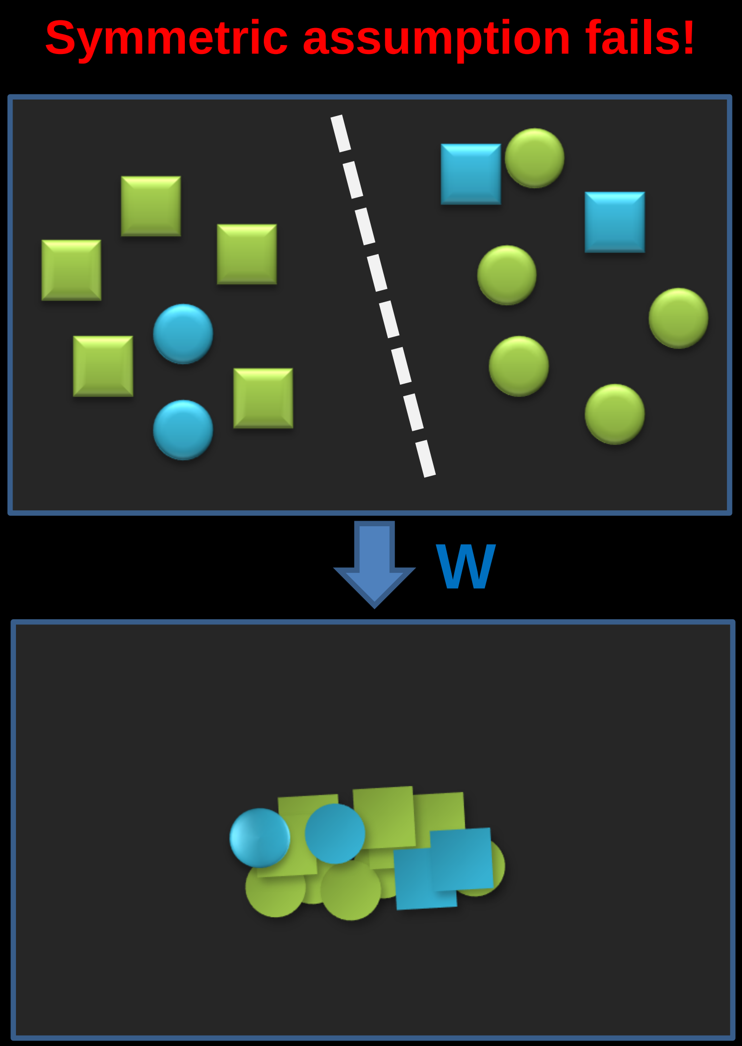
  (W is p.s.d., $W = G^T G$)

- rotation and scaling of all x

✘ cannot handle general shifts

✘ cannot handle new features

How can we learn more general shifts?

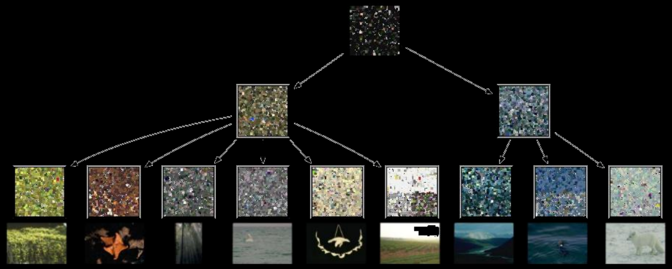


Outline

When statistics of the data change: Cross-domain methods that adapt features

- domain transform [Saenko'10]
- asymmetric transform [Kulis'11]
- manifold walks [Gopalan'11]

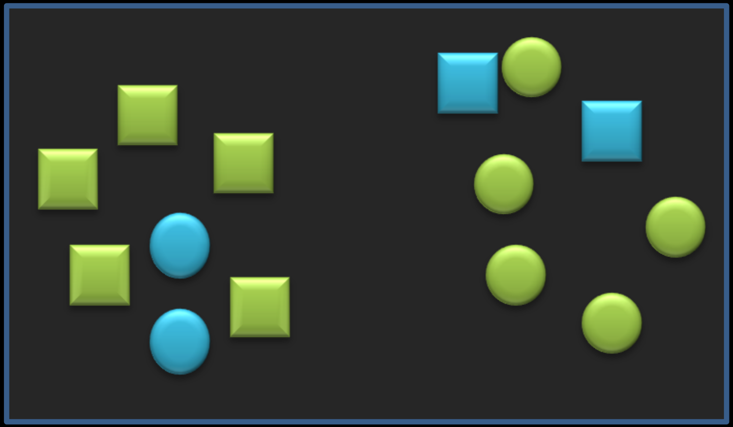
When labels are expensive: Cross-knowledge (same-domain) methods



- sharing features across tasks [Torralba'04], [Quattoni'08]
- visual taxonomies [Bart'08]

[Kulis'11] Improved approach: asymmetric transforms

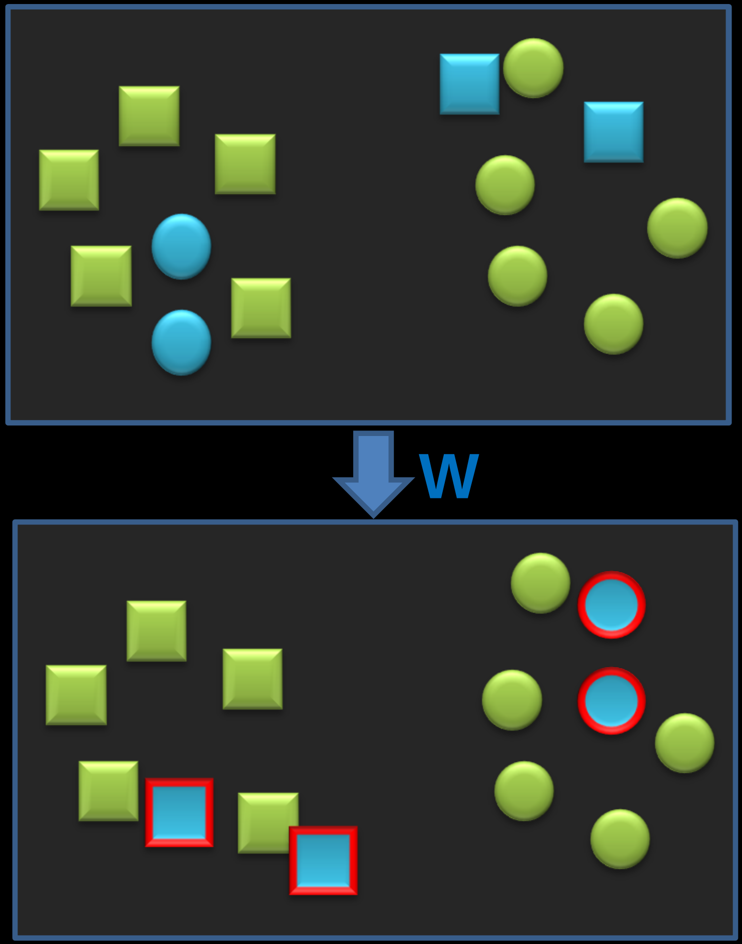
Symmetric assumption fails!



[Kulis'11] Improved approach: asymmetric transforms

- Metric learning model no longer applicable
- We propose to learn asymmetric transforms
  - Map from target to source
  - Handle different dimensions

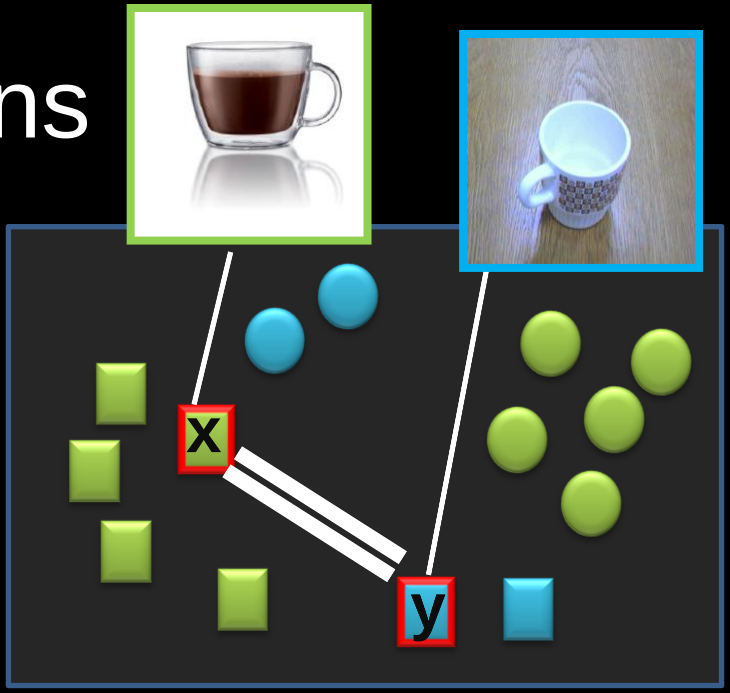
Asymmetric transform (rotation)



[Kulis'11]
Rewrite Loss Functions

Consider inner product between a point *x* from the source and *y* from the target:

$$x^T W y$$



Force to be "large" for similar objects and "small" for dissimilar objects

Use Frobenius norm on W

$$Loss = \begin{cases} (\max(0, \ell - x^T W y))^2 & \text{if } x \text{ and } y \text{ similar,} \\ (\max(0, x^T W y - u))^2 & \text{if } x \text{ and } y \text{ dissimilar.} \end{cases}$$

[Kulis'11] The Model Has Drawbacks

- A linear transformation may be insufficient
- Cost of optimization grows as the product of the dimensionalities of the source and target data

- What to do?

[Kulis'11] Kernelization

- Main idea: run in kernel space
  - Use a non-linear kernel function (e.g., RBF kernel) to learn non-linear transformations in input space
  - Resulting optimization is independent of input dimensionality
  - Additional assumption necessary: regularizer is a *spectral function*

[Kulis'11] Kernelization

$$K_A = X^T X, K_B = Y^T Y$$

Kernel matrices for source and target

Original Transformation Learning Problem

$$\min_W r(W) + \lambda \sum_{i=1}^{m} c_i(X^T W Y)$$

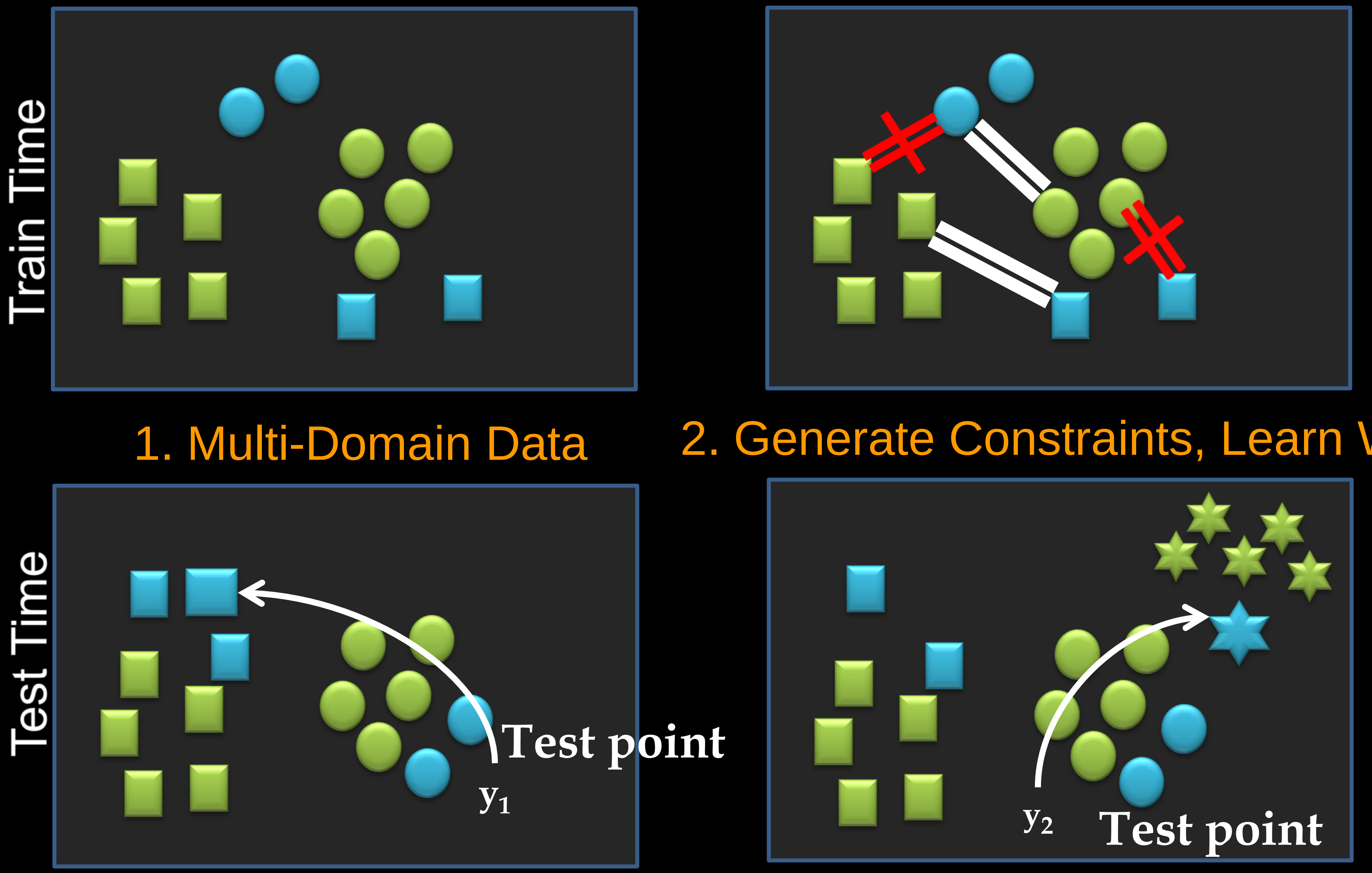
$$\min_L r(L) + \lambda \sum_{i=1}^{m} c_i(K_A^{1/2} L K_B^{1/2})$$

New Kernel Problem

Relationship between original and new problems at optimality

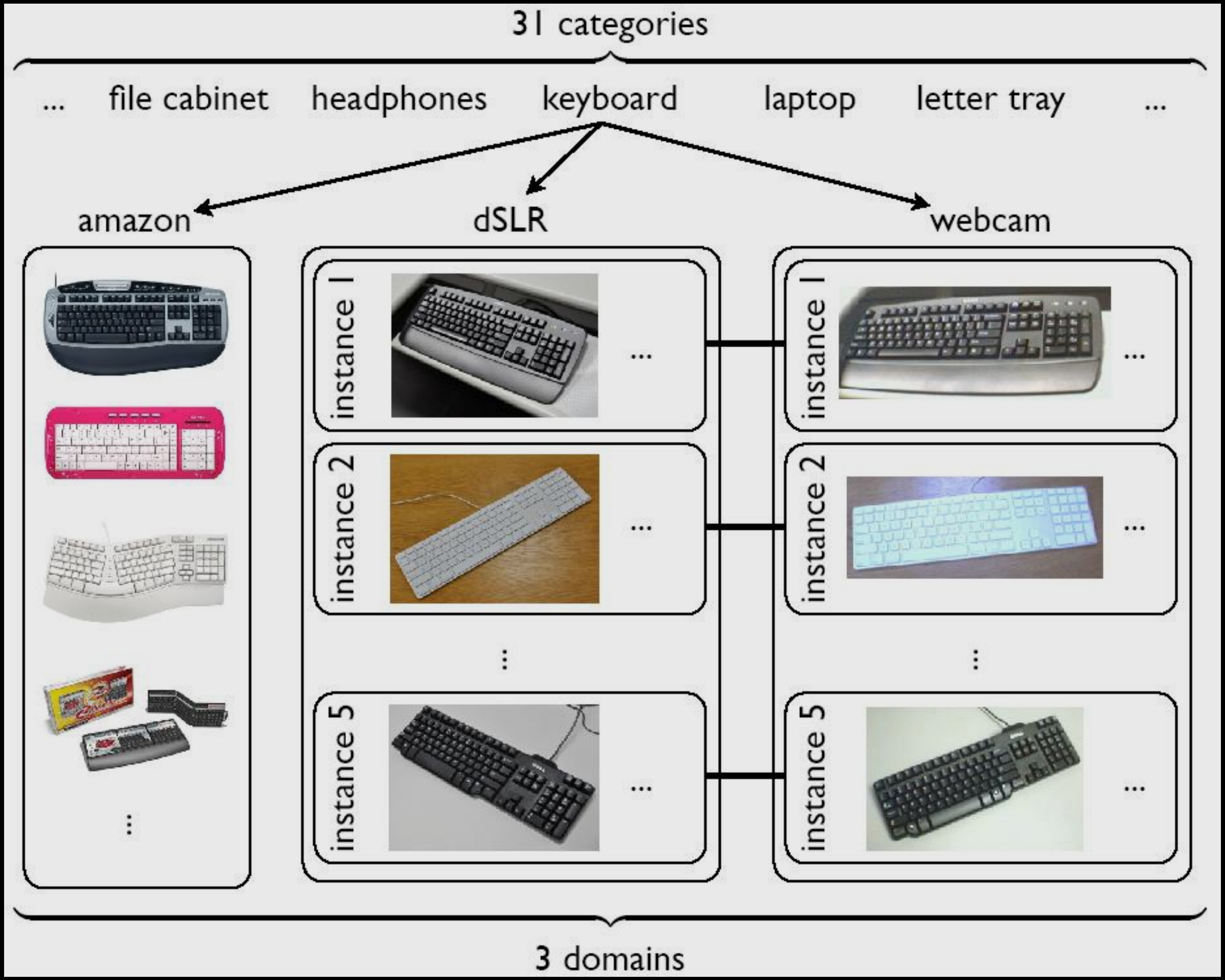
$$W^* = X K_A^{-1/2} L^* K_B^{-1/2} Y^T$$

[Kulis'11] Summary of approach



1. Multi-Domain Data
2. Generate Constraints, Learn W
3. Map via W
4. Apply to New Categories

Multi-domain Office dataset*



*[Saenko et al, ECCV2010]

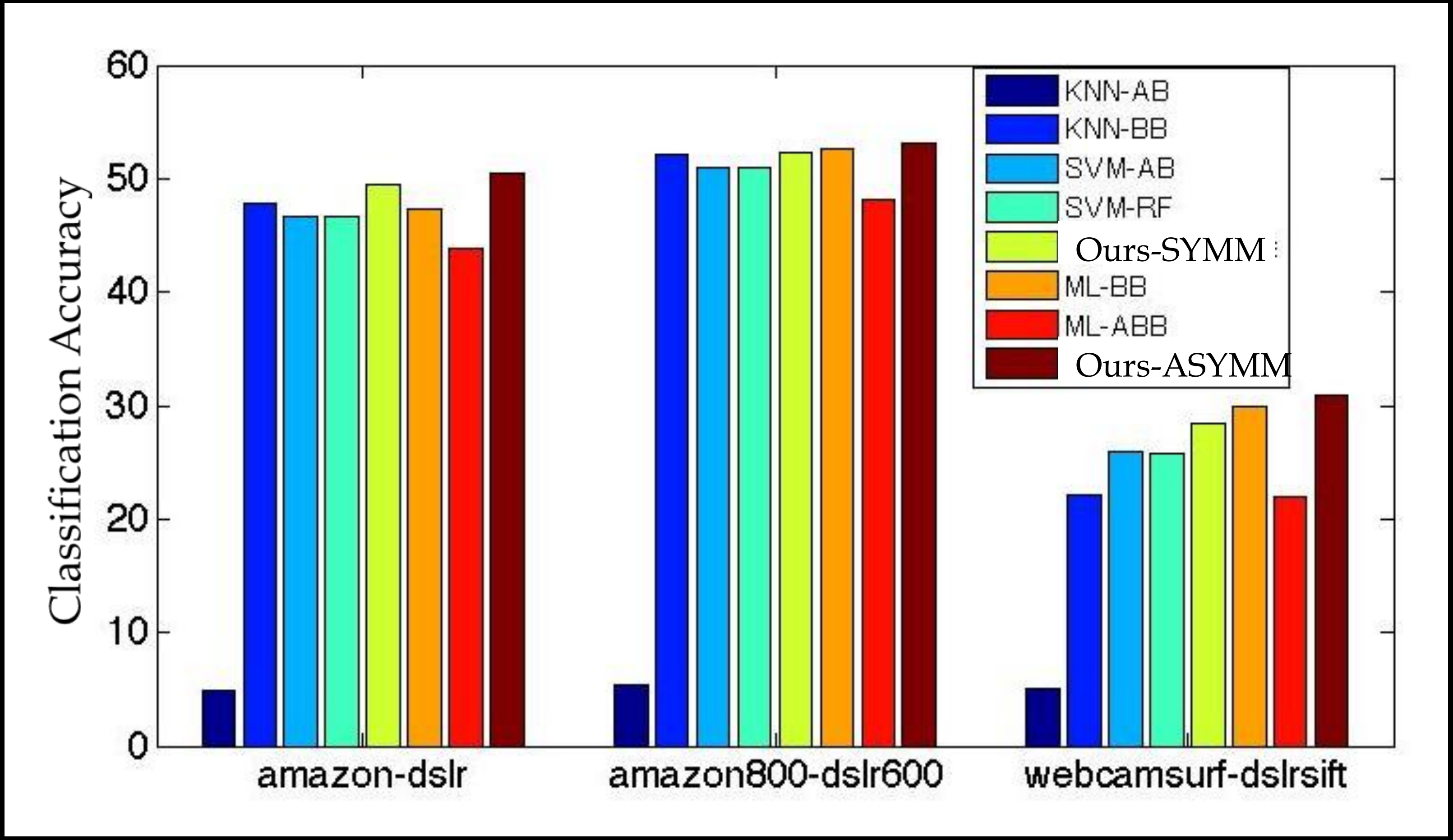
Experimental Setup

- Utilized a standard bag-of-words model using SIFT
  - Adapt from amazon to dslr
- Different dimension of visual word dictionaries of SIFT
  - Adapt from amazon800 to dslr600
- Also utilize different features in the target domain
  - Adapt from webcamSURF to dslrSIFT

Experimental Setup

- Compare to several baselines
  - K-Nearest Neighbors classifier in original feature space
  - SVM in original feature space
  - Metric learning (ITML)
  - nonlinear version of both symm and asymm
- Note: parameter-based methods cannot handle different features!
  - Baseline for comparing such data: KCCA
  - Use same-object correspondences to learn

Same-Category Results on Office



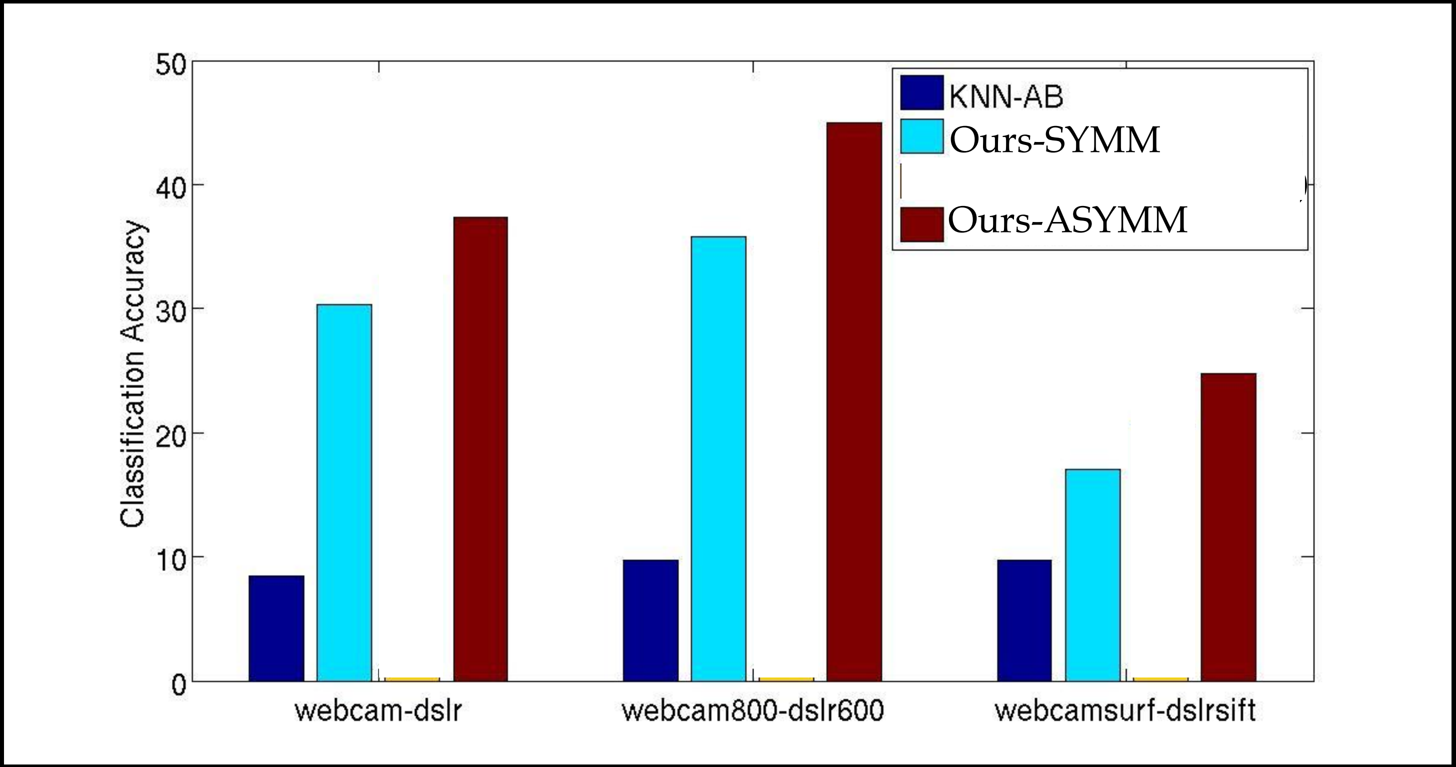
Novel-class experiments

- Test on data from new categories
- Parameter-based methods cannot handle this case
- Compare to linear version of asymmetric method

Experimental Setup

- Utilized a standard bag-of-words model using SIFT
  - Adapt from webcam to dslr
- Different dimension of visual word dictionaries of SIFT
  - Adapt from webcam800 to dslr600
- Also utilize different features in the target domain
  - Adapt from webcamSURF to dslrSIFT

Novel-class experiments on Office



- Test method's ability to transfer domain shift to unseen classes
- Train transform on half of the classes, test on the other half

Extreme shift example

Query from target

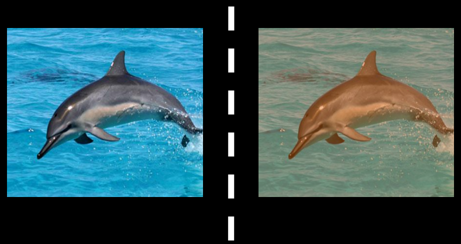
Nearest neighbors in source using KCCA+KNN

Nearest neighbors in source using our method+KNN

Outline

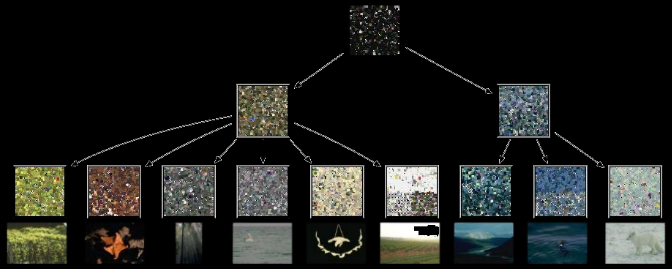
When statistics of the data change: Cross-domain methods that adapt features

- domain transform [Saenko'10]
- asymmetric transform [Kulis'11]
- manifold walks [Gopalan'11]

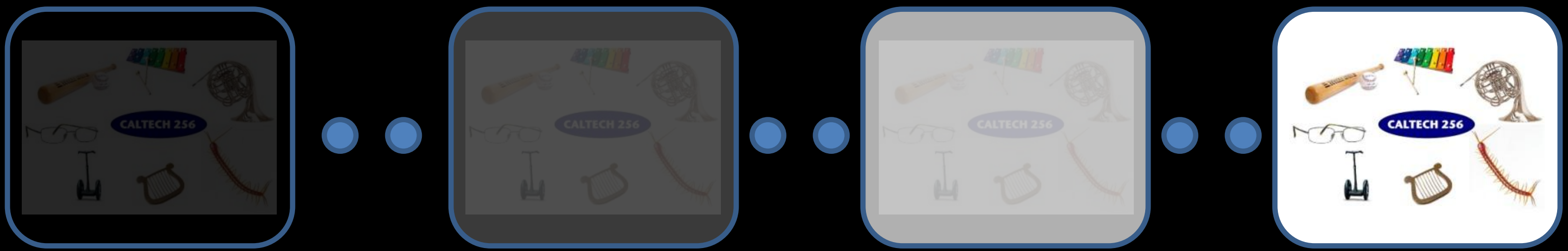


When labels are expensive: Cross-knowledge (same-domain) methods

- sharing features across tasks [Torralba'04], [Quattoni'08]
- visual taxonomies [Bart'08]



[Gopalan'11] Unsupervised Domain Adaptation



Domain 1
(labeled)

Intermediate domains
(Incremental learning)

Domain 2
(unlabeled)

Questions

- How to obtain meaningful intermediate domains?
- How to characterize incremental domain shift information to perform recognition?

slide credit [Gopalan'11]

[Gopalan’11] Unsupervised Domain Adaptation

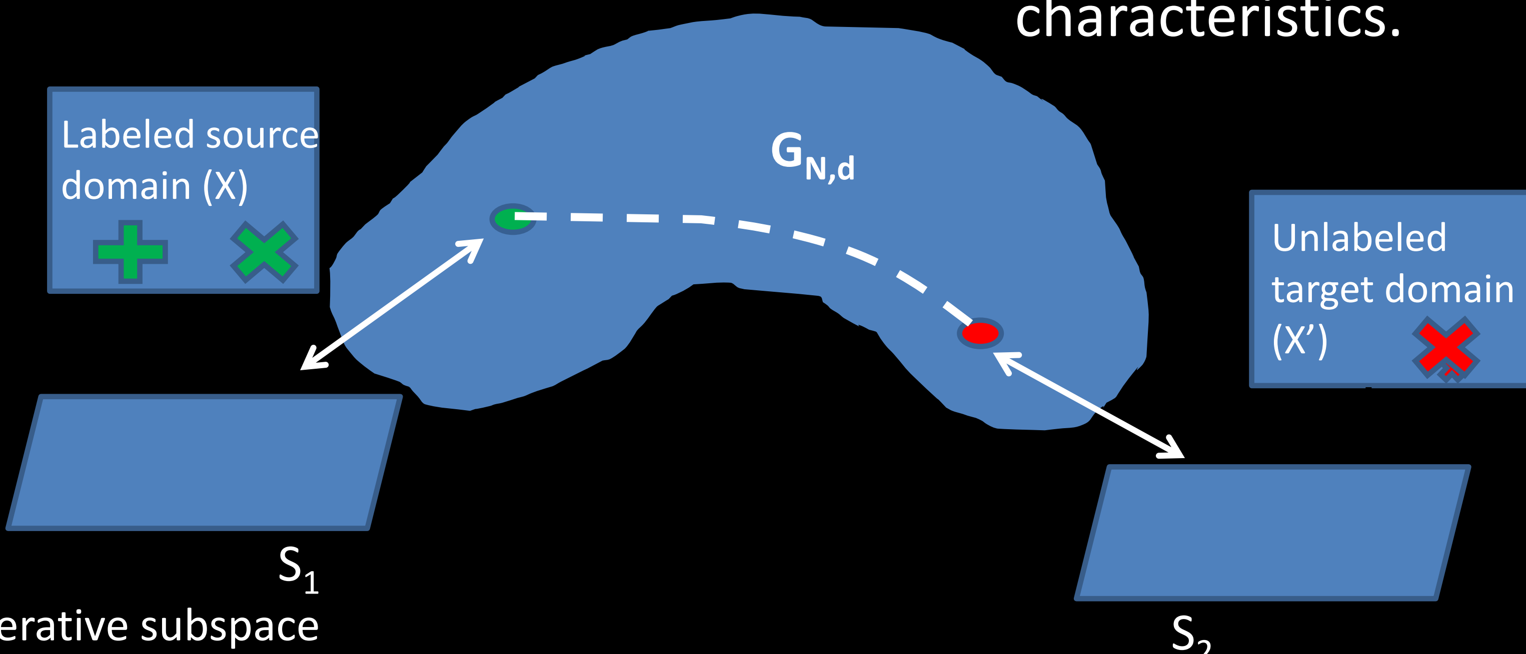
Problem Setting

❖ *d*-dimensional data from *M* object categories.

Domain Representation

N-dimensional generative subspace learned from each domain.

❖ Models *holistic* data characteristics.



Generative subspace from X (no labels)

Generative subspace from X’ (no labels)

slide credit [Gopalan’11]

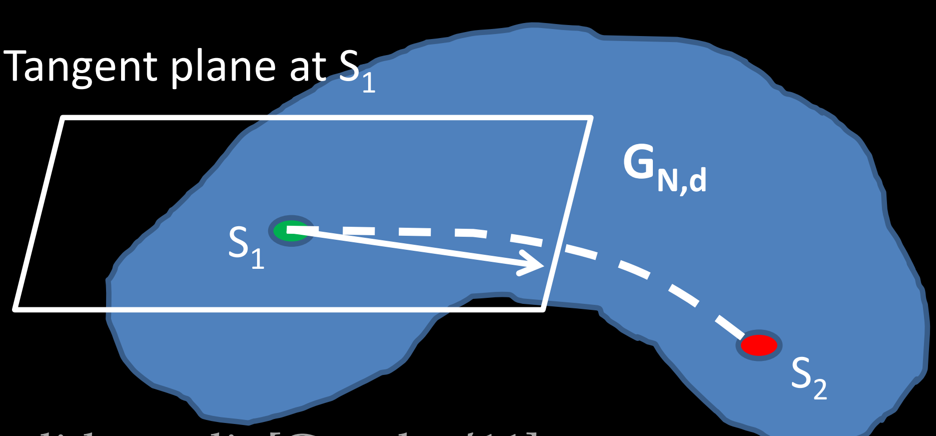
[Gopalan'11] Computations on the Grassmannian

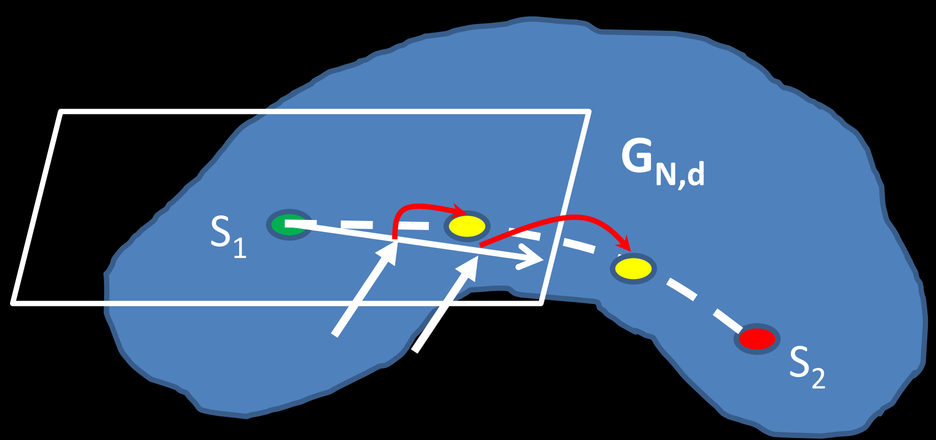
Finding a path between domains

- *Geodesic*
  - Shortest path between points on a manifold
- Computed through the *inverse exponential map*

Generating intermediate subspaces

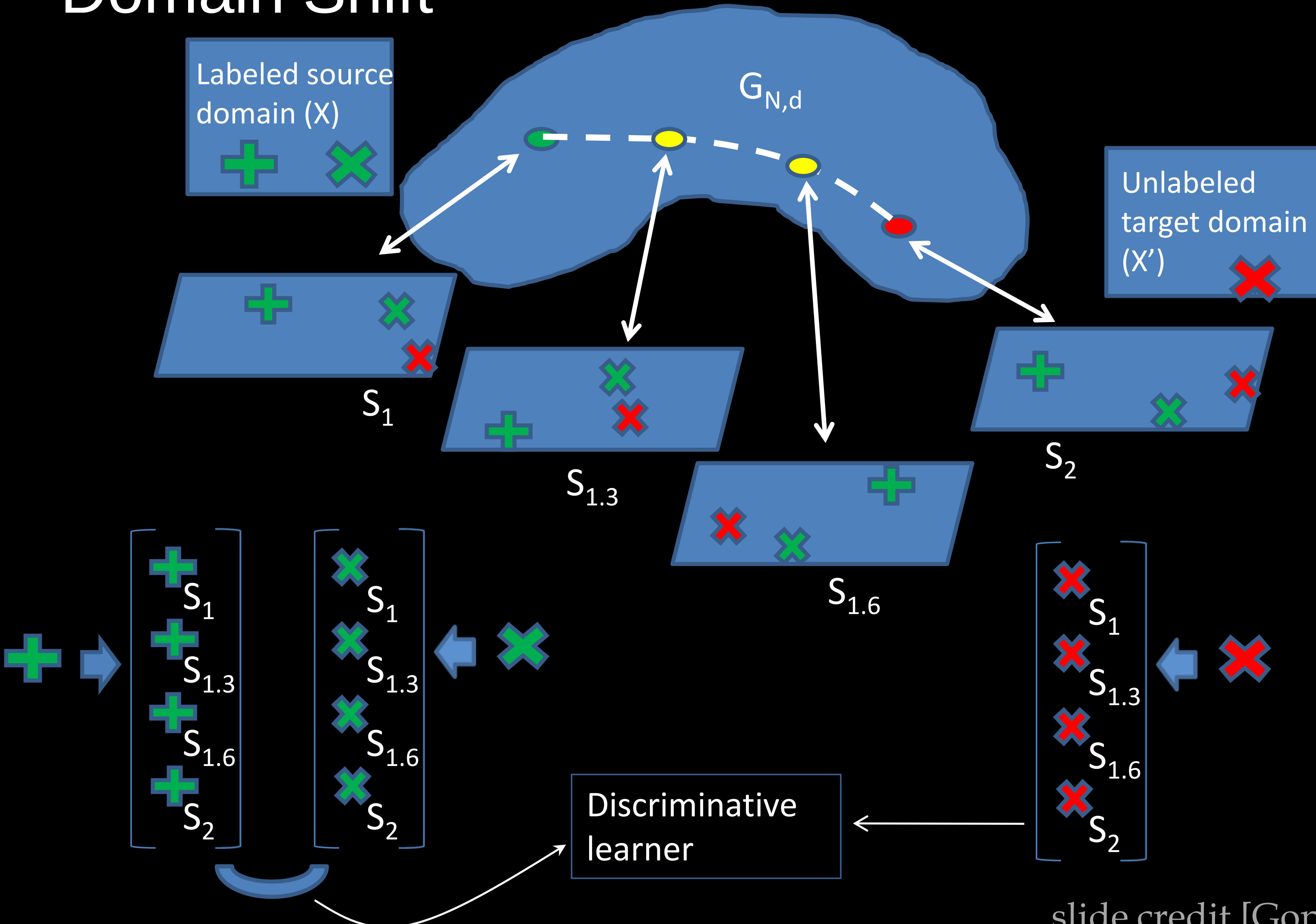
- Sampling `points' along the geodesic
  - *The exponential map*





slide credit [Gopalan'11]

[Gopalan'11] Performing Recognition under Domain Shift

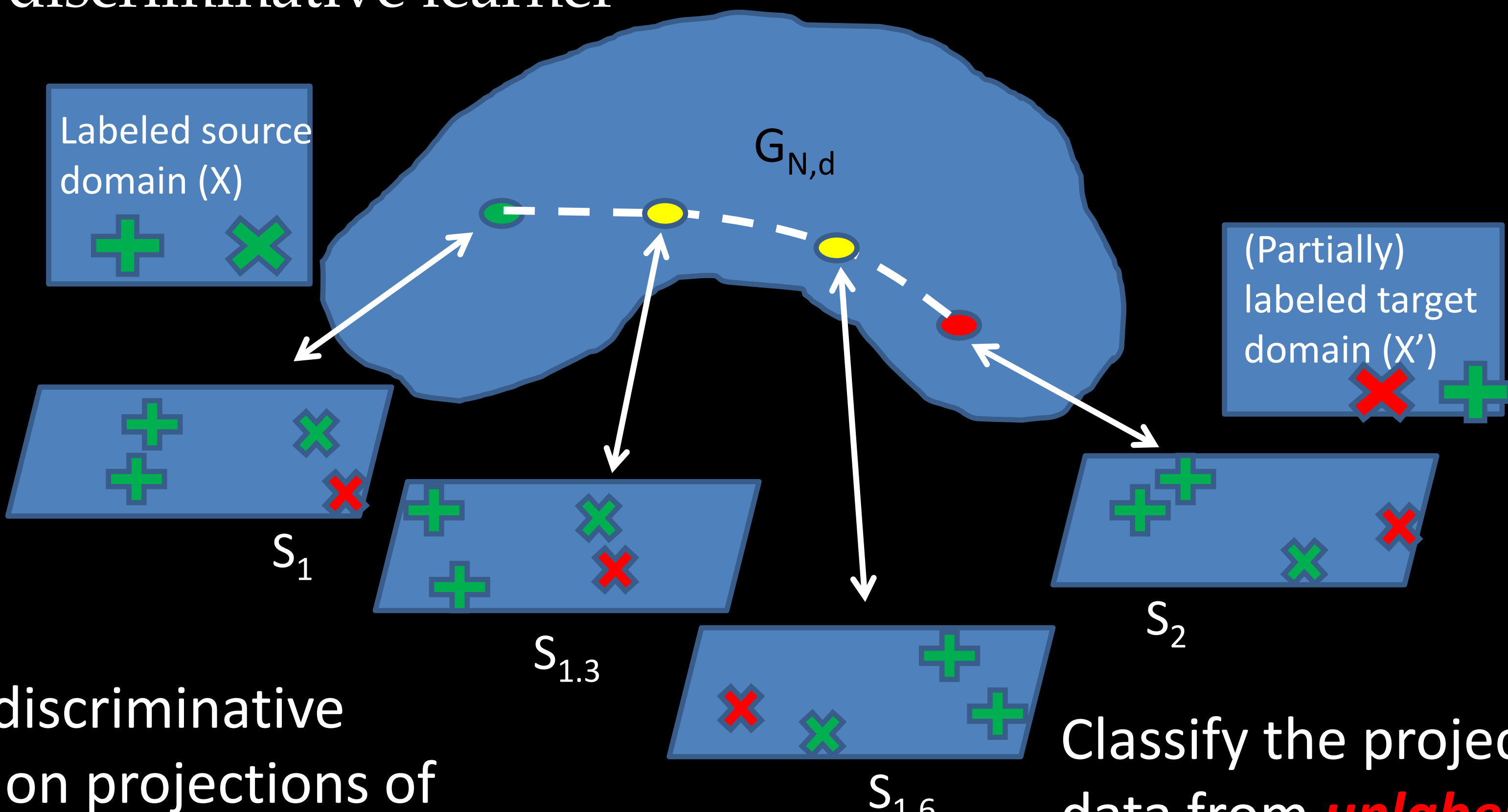


slide credit [Gopalan'11]

[Gopalan'11] Extensions

Semi-supervised domain adaptation

- Use target domain labels in training the discriminative learner



Train a discriminative learner on projections of ***labeled*** data from **X and X'**

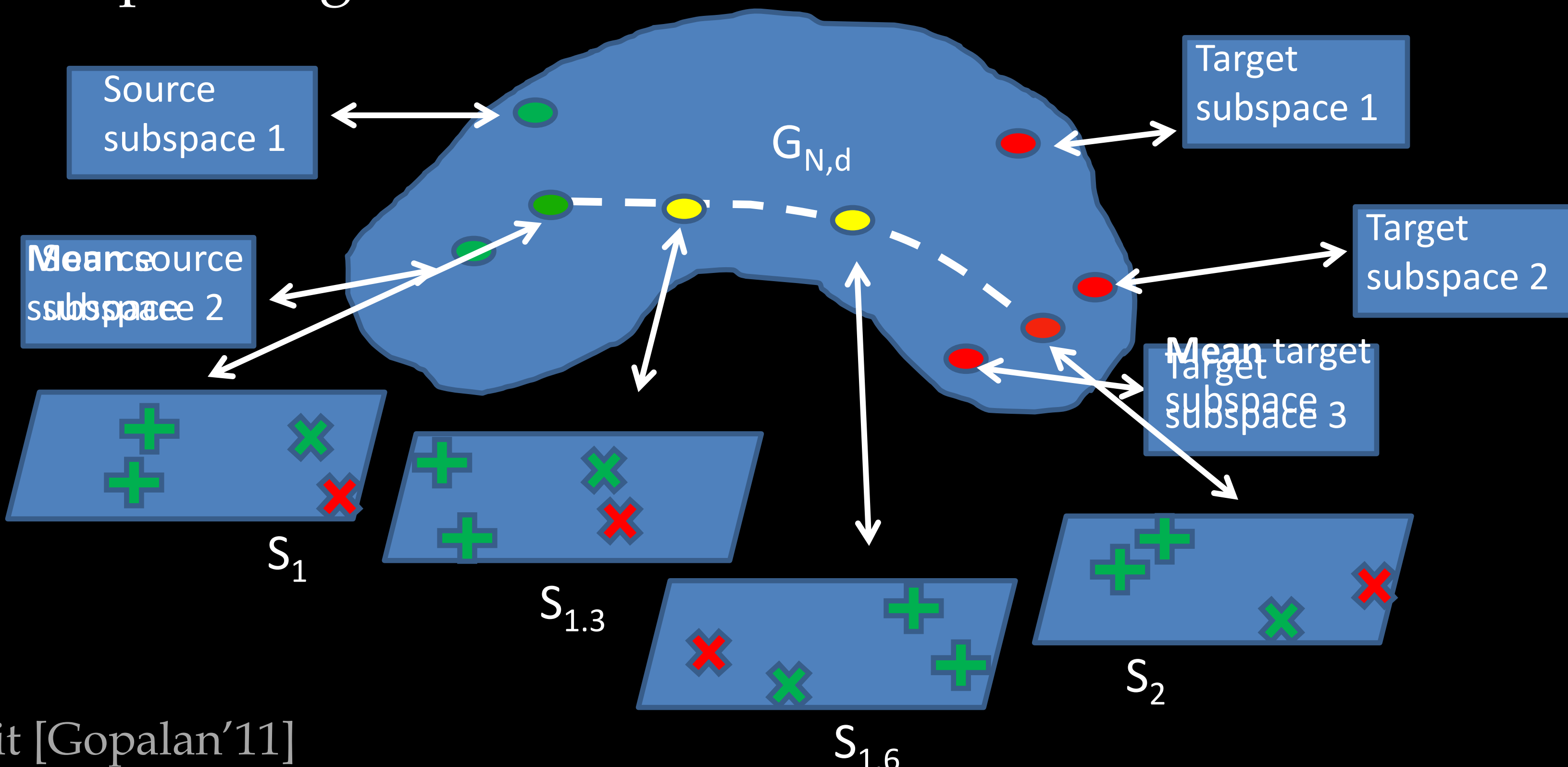
Classify the projected data from ***unlabeled*** **X'**

slide credit [Gopalan'11]

[Gopalan’11] Extensions

Multi-domain adaptation

- k_1 source domains, and k_2 target domains
- find the mean domain for source and target
- sample the geodesic between `mean’ domains



slide credit [Gopalan’11]

Experiment 1: Office domain adaptation

query

retrievals

Domain		Ours	
Source	Target	Un-supervised	Semi-supervised
amazon, dslr	webcam	0.31	0.52
amazon, webcam	dslr	0.25	0.39
dslr, webcam	amazon	0.15	0.28
webcam	amazon, dslr	0.28	0.42
dslr	amazon, webcam	0.35	0.46
amazon	dslr, webcam	0.22	0.32

Multi-domain adaptation

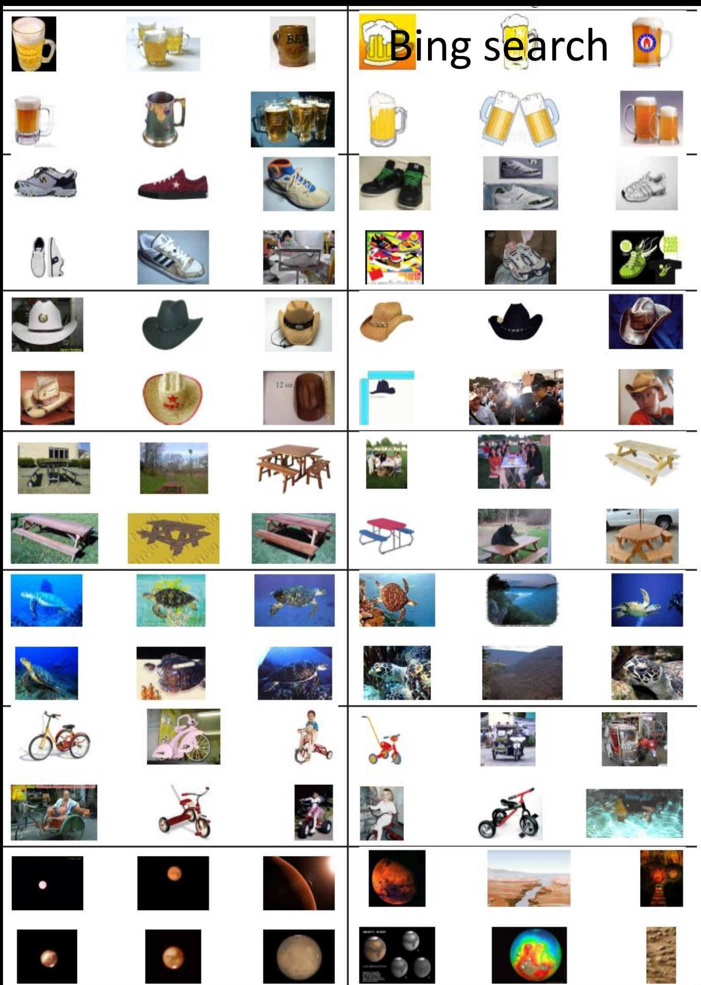
Domain		Metric learning [32] (semi-supervised)		Ours	
Source	Target	asymm	symm	Unsupervised	Semi-supervised
webcam	dslr	0.25	0.27	0.19	0.37
dslr	webcam	0.30	0.31	0.26	0.36
amazon	webcam	0.48	0.44	0.39	0.57

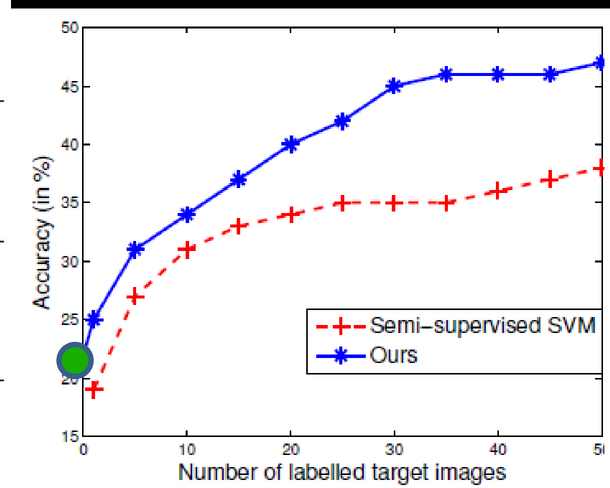
3 domains (webcam, dslr, amazon) – 31 object categories

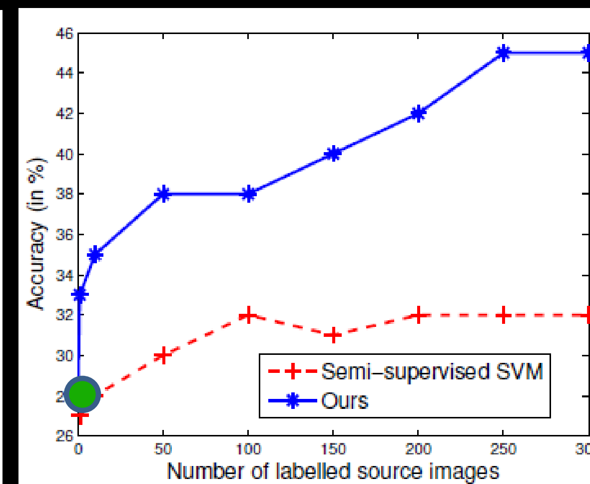
slide credit [Gopalan'11]

Experiment 2: Caltech-256/Bing*

*[Bergamo et al., NIPS 2010]







[1]Target domain: Caltech 256;
Source domain: Images from *Bing* search

[1]L. Torresani, M. Szummer, and A. Fitzgibbon. Efficient object category recognition using classemes. *ECCV 2010*

slide credit [Gopalan'11]

Experiment 3: Sentiment analysis on product reviews*

*[Blitzer et al., ACL 2007]

Domains:
B- books
D- DVD
E- Electronics
K- kitchen appliances

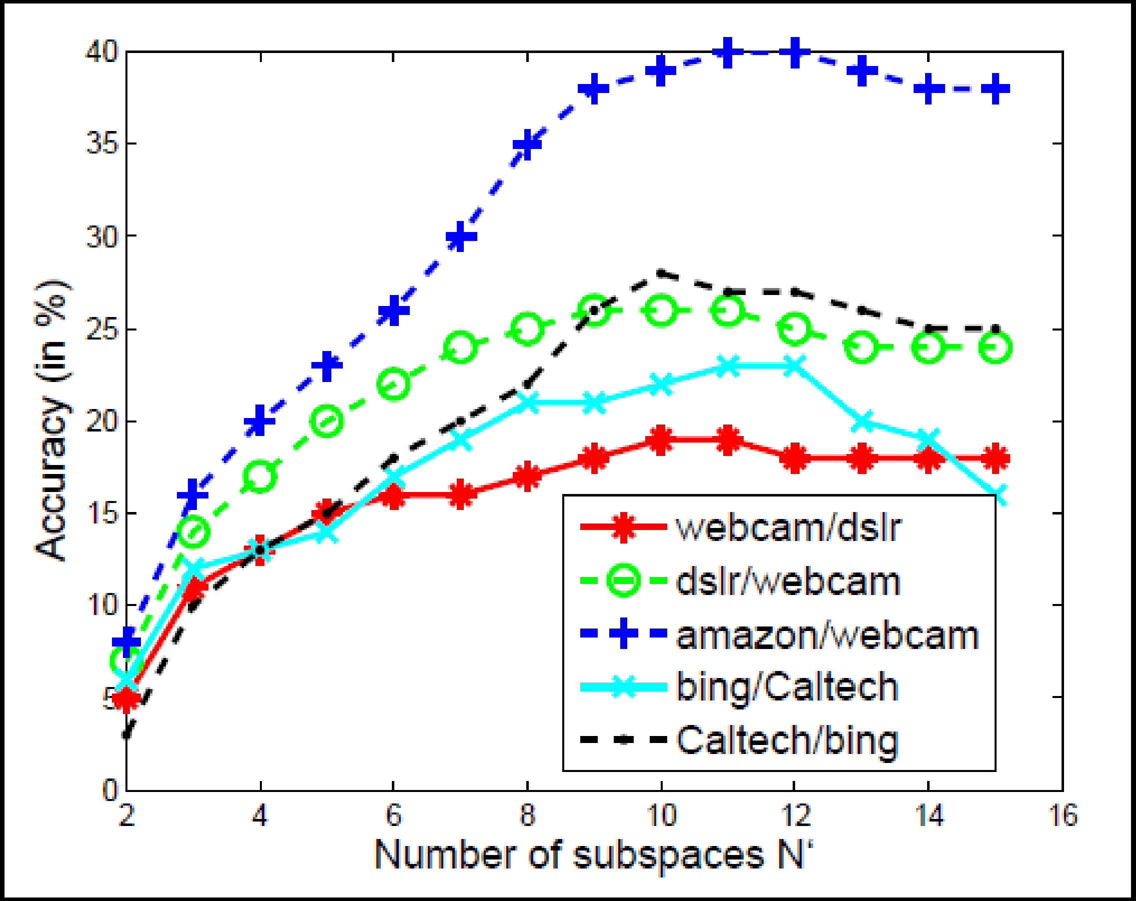
domain\polarity	negative	positive
books	plot <num>_pages predictable reading_this page_<num>	reader grisham engaging must_read fascinating
kitchen	the_plastic poorly_designed leaking awkward_to defective	excellent_product espresso are_perfect years_now a_breeze

Domain		Method Classification (%)		
Target	Source	[10]	[9]	Ours
B	D,E,K	76.8,75.4,66.1	79.7,75.4,68.6	78.2,76.3,74.2
D	B,E,K	74.0,74.3,75.4	75.8,76.2,76.9	76.1,75.8,79.1
E	B,D,K	77.5,74.1,83.7	75.9,74.1,86.8	81.2,76.2,87.6
K	B,D,E	78.7,79.4,84.4	78.9,81.4,85.9	78.1,82.0,89.7

Uses domain-specific knowledge

slide credit [Gopalan'11]

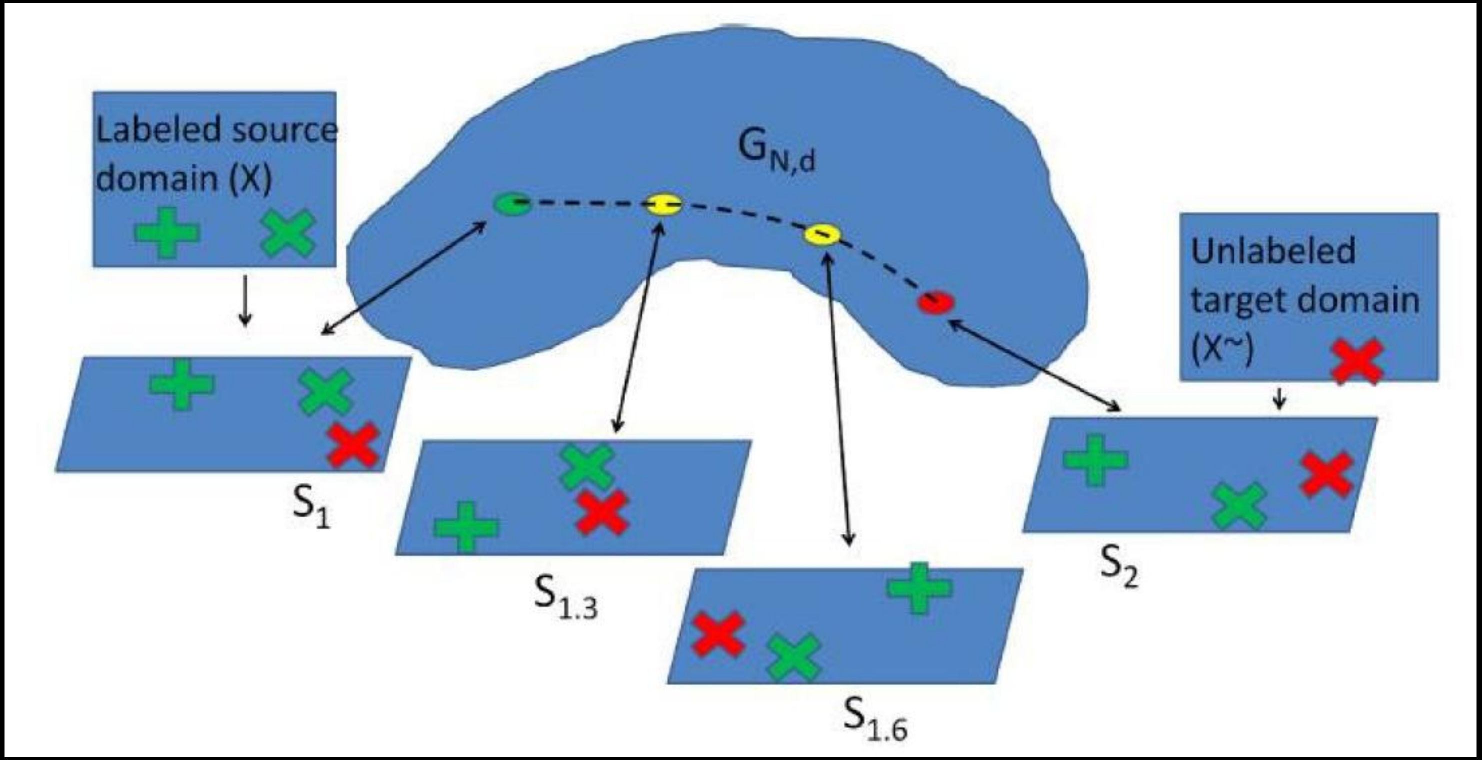
Experiment 4: Information conveyed by *intermediate* domains



slide credit [Gopalan'11]

Discussion of [Gopalan'11]

- A incremental learning-driven approach
  - No assumptions on common domain features/ type of transformation



slide credit [Gopalan'11]

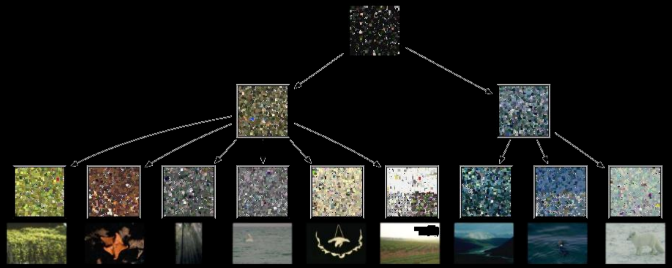
Outline

When statistics of the data change: Cross-domain methods that adapt features

- domain transform [Saenko'10]
- asymmetric transform [Kulis'11]
- manifold walks [Gopalan'11]

When labels are expensive: Cross-knowledge (same-domain) methods

- sharing features across tasks [Torralba'04], [Quattoni'08]
- visual taxonomies [Bart'08]



Transfer learning, or multi-task learning

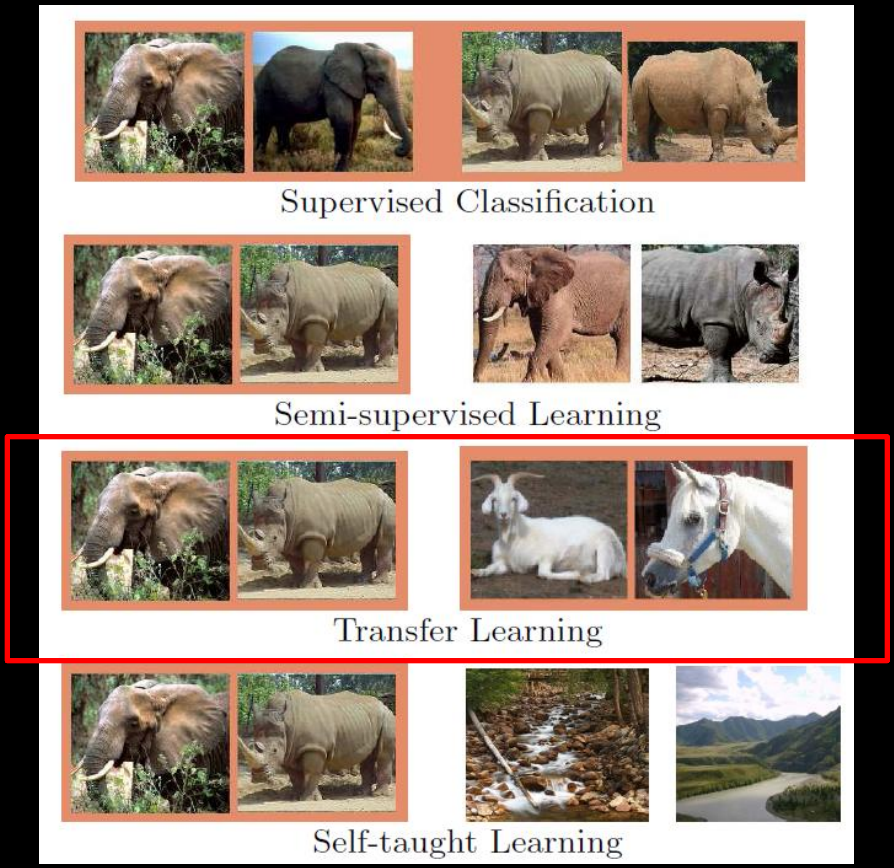


Figure from: [*Raina'07]*

Transfer learning, or multi-task learning

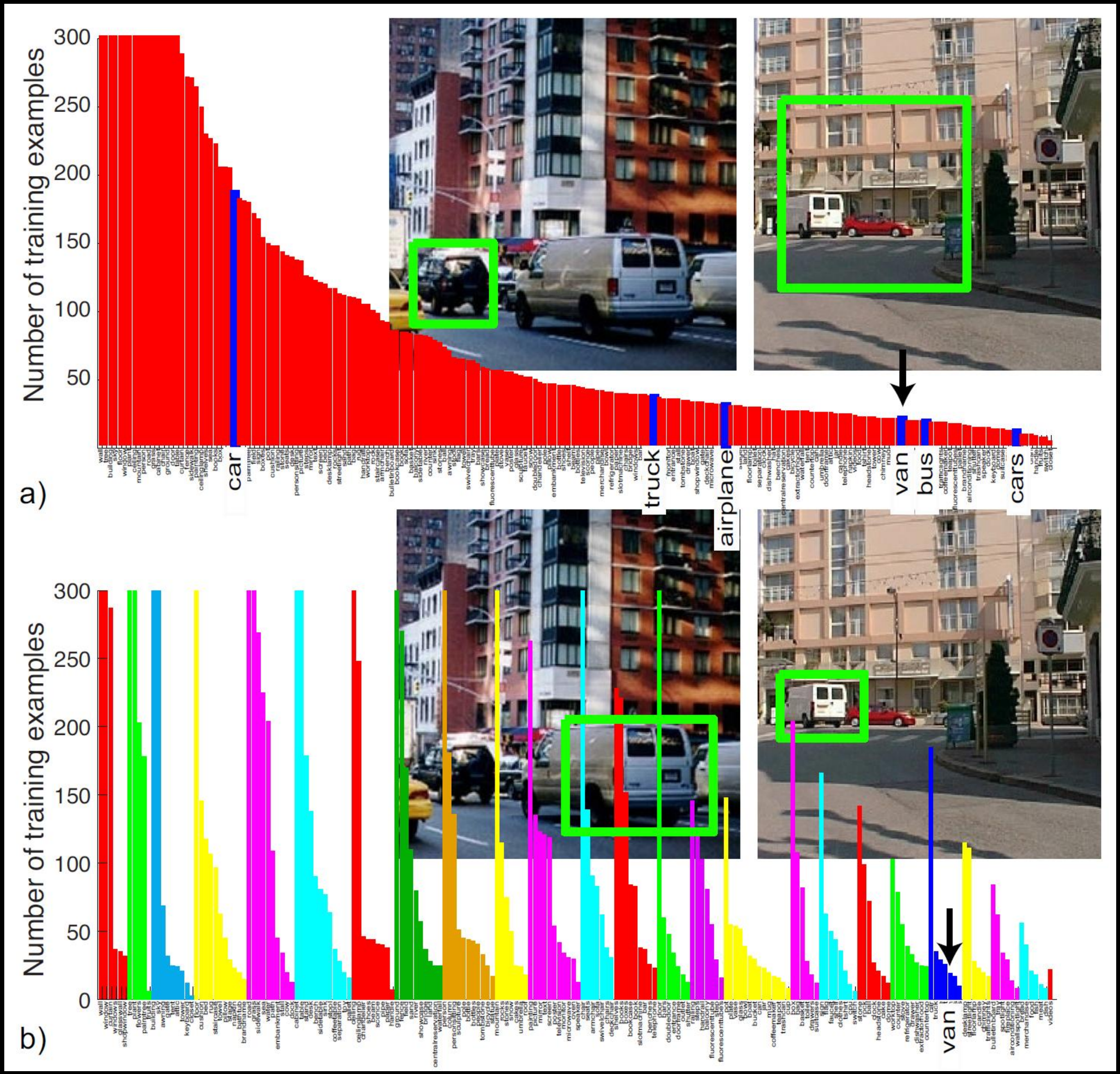


Figure from [Salakhutdinov'11]

Lots of work in this area…

[1] 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.
[2] S. Krempp, D. Geman, and Y. Amit. Sequential learning of reusable parts for object detection. Technical report, CS Johns Hopkins, 2002.
[3] A. Torralba, K. P. Murphy, and W. T. Freeman. Sharing features: efficient boosting procedures for multiclass object detection. In CVPR, 2004.
[4] E. Bart and S. Ullman. Cross-generalization: learning novel classes from a single example by feature replacement. In CVPR, 2005.
[5] A. Opelt, A. Pinz, and A. Zisserman. Incremental learning of object detectors using a visual shape alphabet. In CVPR, 2006.
[6] K. Levi, M. Fink, and Y. Weiss. Learning from a small number of training examples by exploiting object categories. In Workshop of Learning in Computer Vision, 2004.
[7] A. Quattoni, M. Collins, and T.J. Darrell. Transfer learning for image classification with sparse prototype representations. In CVPR, 2008.
[8] C.H. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In CVPR, 2009.
[9] E. Miller, N. Matsakis, and P. Viola. Learning from one example through shared densities on transforms. In CVPR, 2000.
[10] L. Fei-Fei, R. Fergus, and P. Perona. A bayesian approach to unsupervised one-shot learning of object categories. In ICCV, 2003.
[11] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. In IEEE. Workshop on GMBV, 2004.
[12] E. Sudderth, A. Torralba,W. T. Freeman, andW.Willsky. Learning hierarchical models of scenes, objects, and parts. In ICCV, 2005.
[13] J. Sivic, B.C. Russell, A. Zisserman, W.T. Freeman, and A.A. Efros. Unsupervised discovery of visual object class hierarchies. In CVPR, 2008.
[14] E. Bart, I. Porteous, P. Perona, and M. Welling. Unsupervised learning of visual taxonomies. In CVPR, 2008…

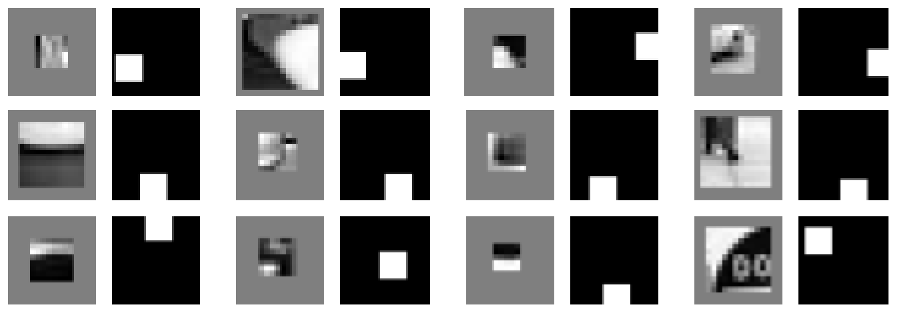
…I will cover three representative papers

[1] 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.
[2] S. Krempp, D. Geman, and Y. Amit. Sequential learning of reusable parts for object detection. Technical report, CS Johns Hopkins, 2002.
[3] A. Torralba, K. P. Murphy, and W. T. Freeman. Sharing features: efficient boosting procedures for multiclass object detection. In CVPR, 2004.
[4] E. Bart and S. Ullman. Cross-generalization: learning novel classes from a single example by feature replacement. In CVPR, 2005.
[5] A. Opelt, A. Pinz, and A. Zisserman. Incremental learning of object detectors using a visual shape alphabet. In CVPR, 2006.
[6] K. Levi, M. Fink, and Y. Weiss. Learning from a small number of training examples by exploiting object categories. In Workshop of Learning in Computer Vision, 2004.
[7] A. Quattoni, M. Collins, and T.J. Darrell. Transfer learning for image classification with sparse prototype representations. In CVPR, 2008.
[8] C.H. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In CVPR, 2009.
[9] E. Miller, N. Matsakis, and P. Viola. Learning from one example through shared densities on transforms. In CVPR, 2000.
[10] L. Fei-Fei, R. Fergus, and P. Perona. A bayesian approach to unsupervised one-shot learning of object categories. In ICCV, 2003.
[11] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. In IEEE. Workshop on GMBV, 2004.
[12] E. Sudderth, A. Torralba,W. T. Freeman, and W.Willsky. Learning hierarchical models of scenes, objects, and parts. In ICCV, 2005.
[13] J. Sivic, B.C. Russell, A. Zisserman, W.T. Freeman, and A.A. Efros. Unsupervised discovery of visual object class hierarchies. In CVPR, 2008.
[14] E. Bart, I. Porteous, P. Perona, and M. Welling. Unsupervised learning of visual taxonomies. In CVPR, 2008…

[Torralba'04] Sharing features for object detection

- Task: object detection
- boosting detector
  - combines multiple features (weak learners)

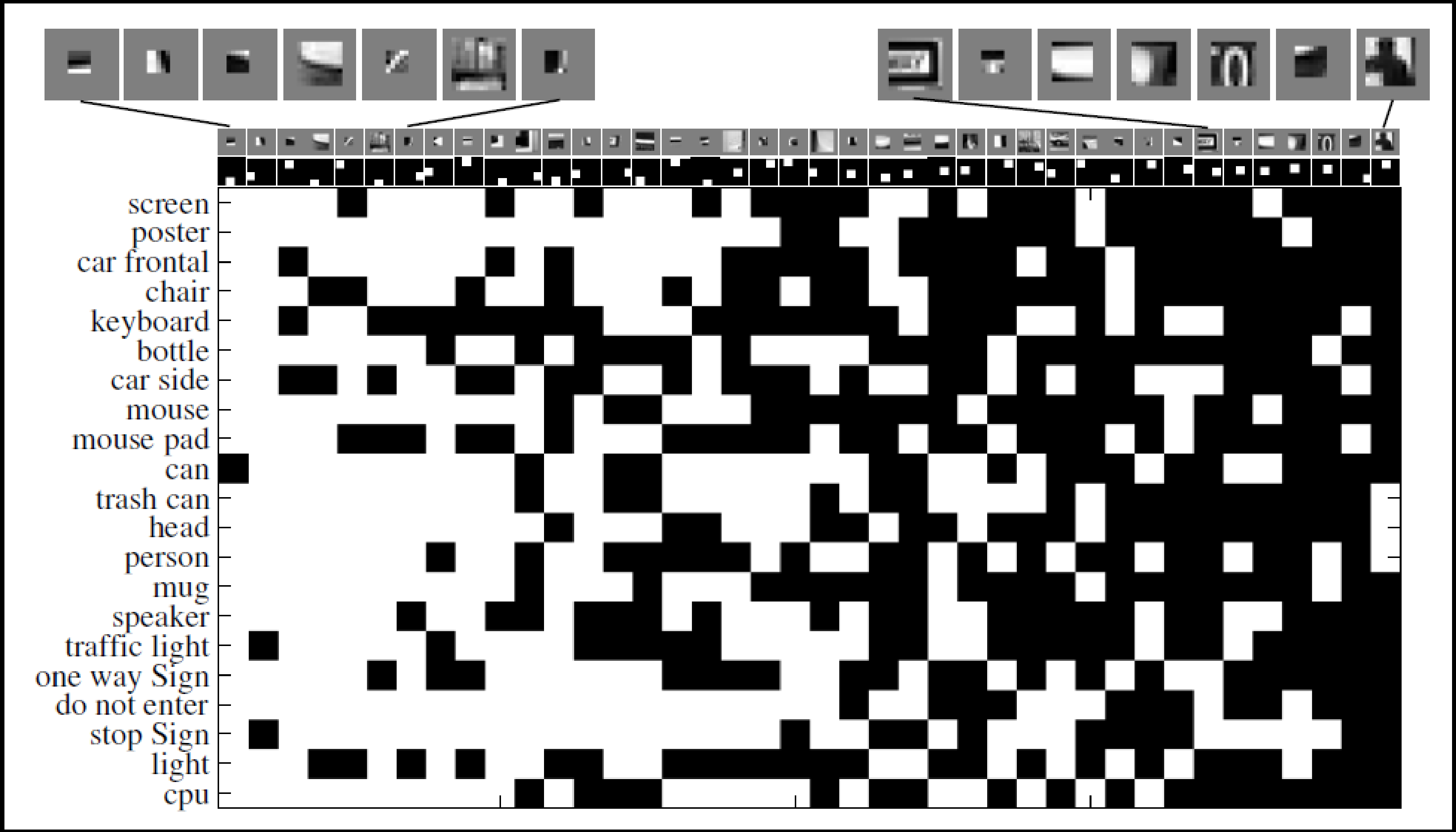
Idea: Share features across tasks (objects)



features

[Torralba'04] Sharing features for object detection

- Shared features across object classes



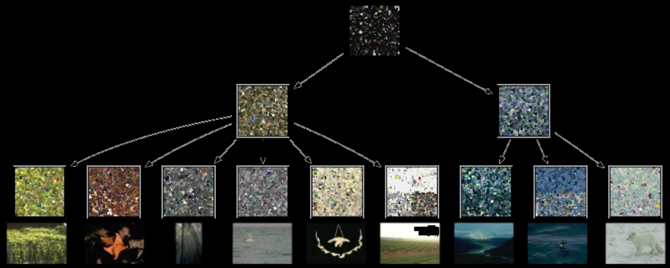
Outline

When statistics of the data change: Cross-domain methods that adapt features

- domain transform [Saenko'10]
- asymmetric transform [Kulis'11]
- manifold walks [Gopalan'11]

When labels are expensive: Cross-knowledge (same-domain) methods

- sharing features across tasks [Torralba'04], [Quattoni'08]
- visual taxonomies [Bart'08]



[Quattoni'08] Sparse Prototypes

- Task: image-based news topic prediction

Related news topics

SuperBowl

Sharon

Danish Cartoons

Figure Skating

Australian Open

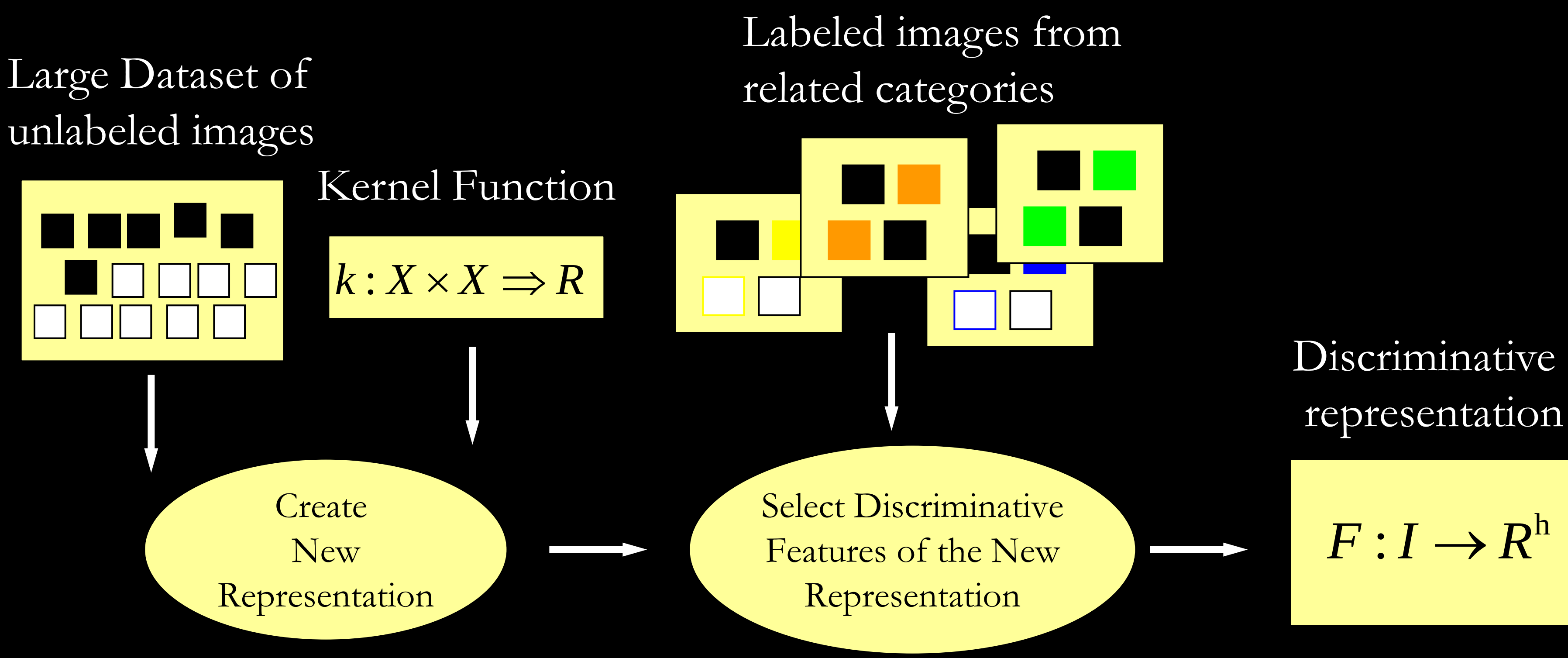
labeled

Novel news topic

Academy Awards

few labels

[Quattoni'08] Sparse Prototypes



Slide credit: A. Quattoni

[Quattoni'08] Sparse Prototypes

Preliminaries

Input 1: Unlabeled dataset
$U = \{x_1, x_2, \ldots, x_p\}$ for $x_i \in \mathcal{X}$ (e.g, $\mathcal{X} = \mathbb{R}^d$)
Input 2: Collection of related problems
$C = \{T_1, \ldots, T_m\}$ where
$T_k = \{(x_1^k, y_1^k), (x_2^k, y_2^k), \ldots, (x_{n_k}^k, y_{n_k}^k)\}$
for $x \in \mathcal{X}$ and $y \in \{+1, -1\}$
Input 3: Kernel function
$k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$
Input 4: Threshold θ
Input 5: Regularization constants λ_k, for $k = 1 : m$

Slide credit: A. Quattoni

[Quattoni'08] Sparse Prototypes

Step 1: Compute prototypes using unlabeled data

- Compute the kernel matrix for all unlabeled points :
  $K_{ij} = k(x_i, x_j)$ for $x_i \in U$, $x_j \in U$

- Compute eigenvectors of of K by performing SVD :
  Compute a projection matrix A of dimension $p \times p$ by taking the eigenvectors of K; where each column of A corresponds to an eigenvector.

- Project all points x_i^k in C to the prototype space:
  $z(x_i^k) = A^\top \varphi(x_i^k)$
  where
  $\varphi(x) = [k(x, x_1), \ldots, k(x, x_p)]^\top, \quad x_i \in U$

Slide credit: A. Quattoni

[Quattoni'08] Sparse Prototypes

Step 2: Use related tasks to select a sparse set of the most discriminative prototypes

Let W be a $p \times m$ matrix where W_{jk} corresponds to the j-th coefficient of the k-th problem.

- Choose the optimal matrix W^* to be:

  $$\min_{W,\varepsilon} \sum_{k=1}^{m} \lambda_k \sum_{i=1}^{n_k} \varepsilon_i^k + \sum_{j=1}^{p} \max_k |W_{jk}|$$

  $s.t.$ for $k = 1 : m$ and $i = 1 : n_k$

  $$y_i^k \mathbf{w}_k^\top z(x_i^k) \geq 1 - \varepsilon_i^k$$

  $$\varepsilon_i^k \geq 0$$

  where $\mathbf{w}_k$ is the k-th column of W, corresponding to the parameters for problem k.

Slide credit: A. Quattoni

[Quattoni'08] Sparse Prototypes

Step 3: Create new representation based on kernel distances to selected prototypes

- Define the set of relevant prototypes to be:
  $$R = \{r : \max_k |W^*_{rk}| > \theta\}$$
- Create projection matrix B by taking all the columns of A corresponding to the indexes in R. B is then a $p \times h$ matrix, where $h = |R|$.
- Return the representation given by:
  $$v(x) = B^\top \varphi(x)$$

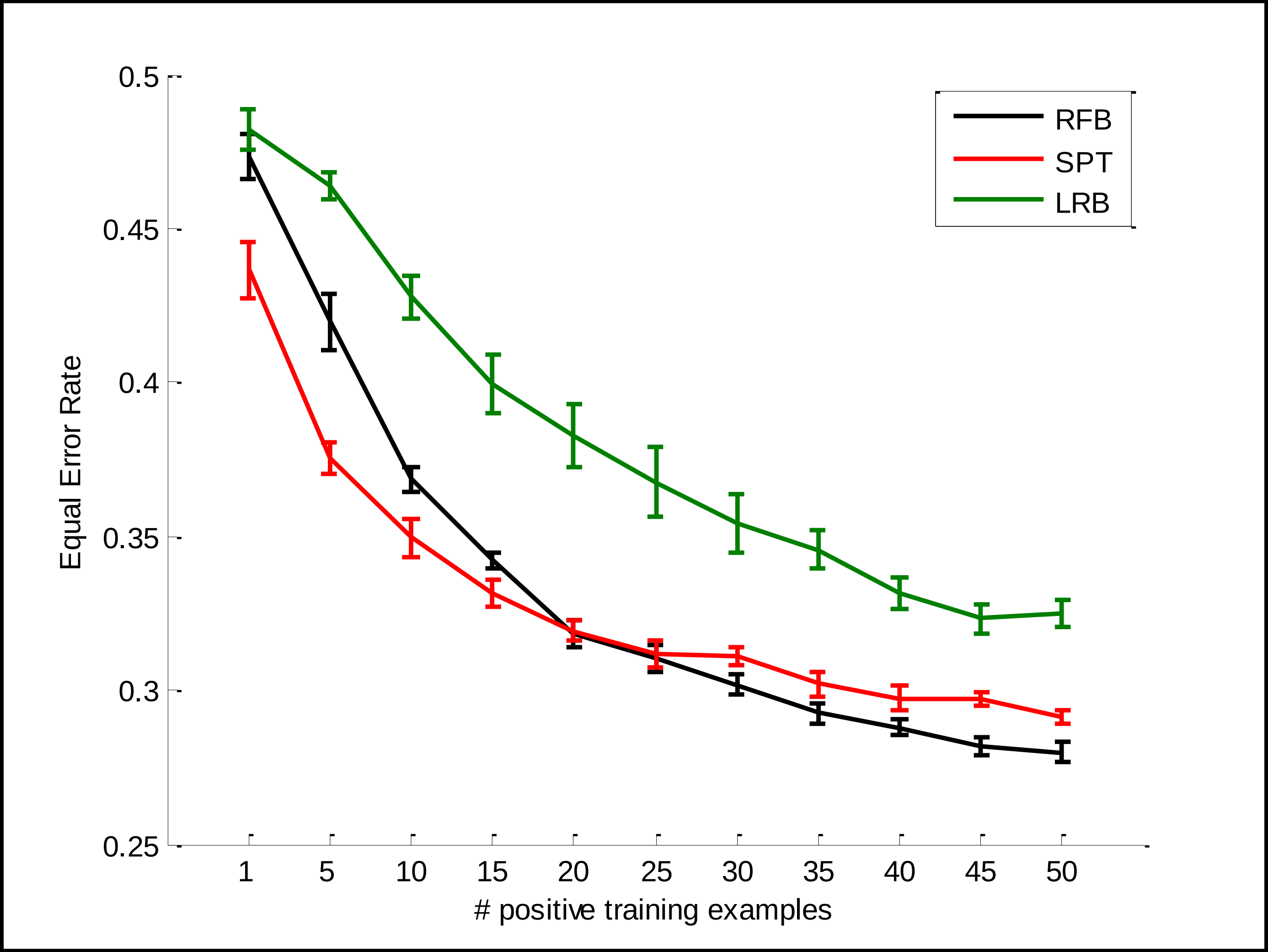
Slide credit: A. Quattoni

Experiments: News topic classification

- Three models, all linear SVMs
- Baseline model (**RFB**):
  - Uses raw representation.
- Low Rank baseline (**LRB**):
  - Uses as a representation:
    where **Q** is consists of the **h** highest eigenvectors of the matrix **A** computed in the first step of the algorithm
- Sparse Transfer Model (**SPT**)
  - Uses Representation computed by our algorithm
- For both **LRB** and **SPT** we used and **RBF** kernel when computing the Representation from unlabeled data.

Slide credit: A. Quattoni

Experiments: News topic classification

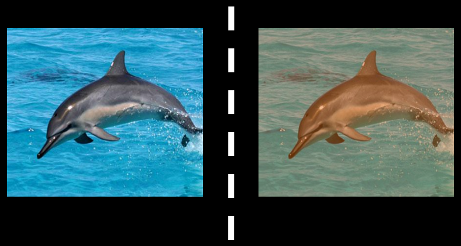


Slide credit: A. Quattoni

Outline

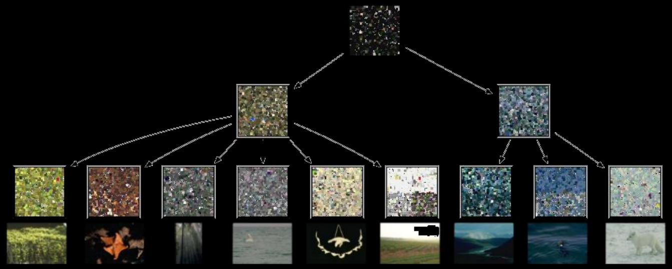
When statistics of the data change: Cross-domain methods that adapt features

- domain transform [Saenko’10]
- asymmetric transform [Kulis’11]
- manifold walks [Gopalan’11]



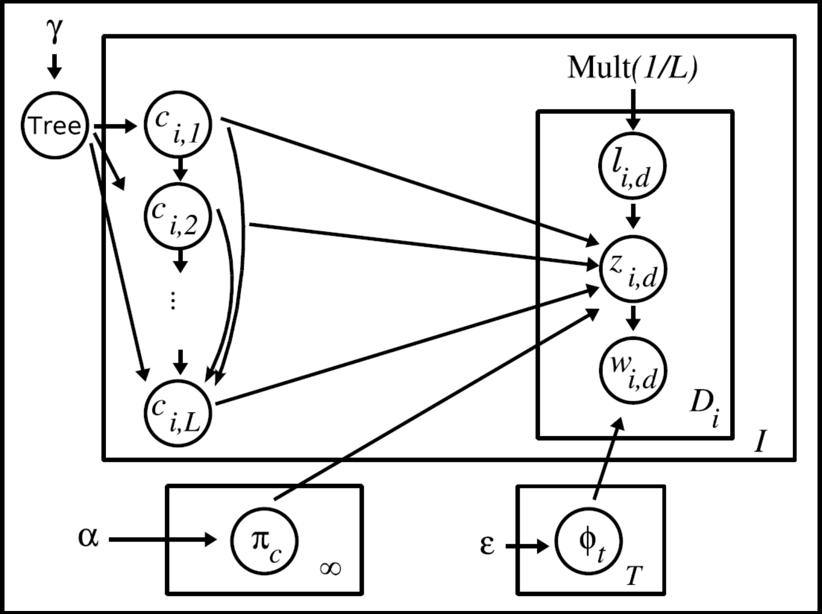
When labels are expensive: Cross-knowledge (same-domain) methods

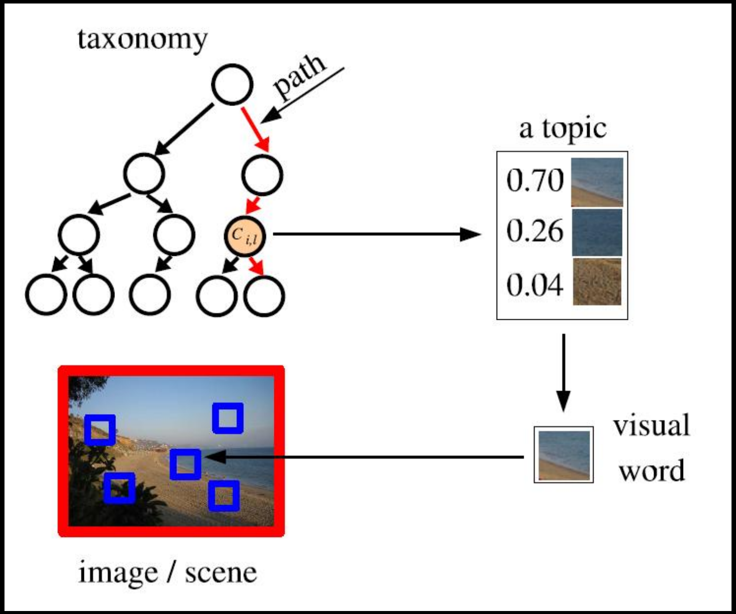
- sharing features across tasks [Torralba’04], [Quattoni’08]
- visual taxonomies [Bart’08]



[Bart'08] Visual taxonomies

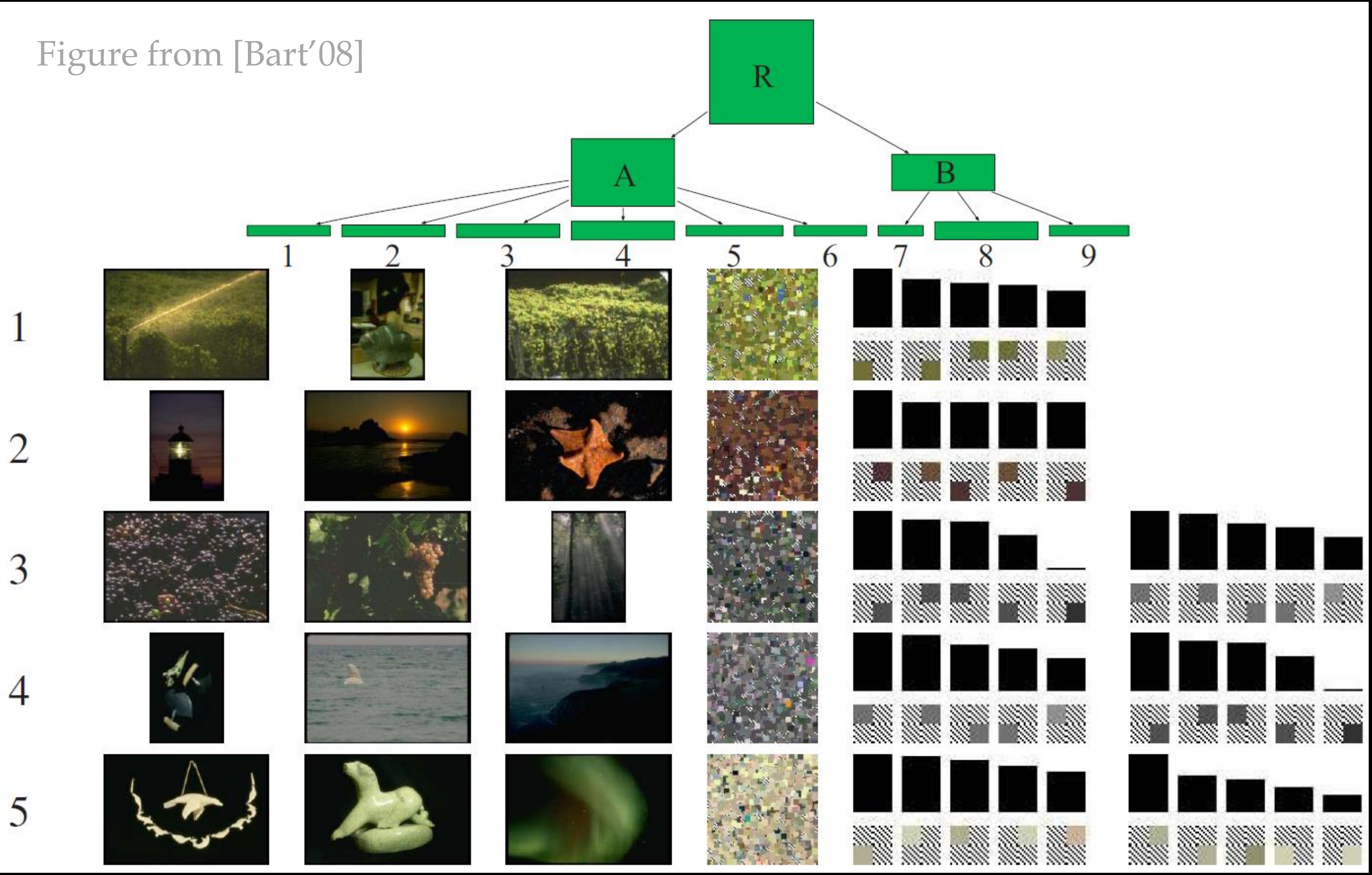
- Discover shared hierarchy of images
- Bayesian nonparametric model





Figures from [Bart'08]

Taxonomy learned on 300 Corel images



Summary

- Covered methods that learn new representations to transfer knowledge
  - from source to target domain
  - from related tasks to novel tasks
  - maximize use of available data!
- Representations can capture
  - change in feature distribution between data sources
  - change in feature type and dimensions
  - common features of related tasks

Questions?

References

A. Torralba, K. P. Murphy, and W. T. Freeman. Sharing features: efficient boosting procedures for multiclass object detection. In CVPR, 2004.

H. Daume III. Frustratingly easy domain adaptation. In ACL, 2007.

R. Raina, A. Battle, H. Lee, B. Packer and A.Y. Ng. Self-taught learning: Transfer learning from unlabeled data. In ICML, 2007.

J. Yang, R. Yan, and A. G. Hauptmann. Cross-domain video concept detection using adaptive svms. ACM Multimedia, 2007.

E. Bart, I. Porteous, P. Perona, and M. Welling. Unsupervised learning of visual taxonomies. In CVPR, 2008.

W. Dai, Y. Chen, G.-R. Xue, Q. Yang, and Y. Yu. Translated learning: Transfer learning across different feature spaces. In NIPS, 2008.

A. Farhadi and M. K. Tabrizi. Learning to recognize activities from the wrong view point. In ECCV, 2008.

References (cont’d)

A. Quattoni, M. Collins, and T.J. Darrell. Transfer learning for image classification with sparse prototype representations. In CVPR, 2008.

L. Duan, I. W. Tsang, D. Xu, and S. J. Maybank. Domain transfer svm for video concept detection. In CVPR, 2009.

L. Duan, D. Xu, I. Tsang, and J. Luo. Visual event recognition in videos by learning from web data. In Proc. CVPR, 2010.

K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In ECCV, 2010.

B. Kulis, K. Saenko and T. Darrell. Adapting visual category models to new domains. In CVPR, 2011.

R. Salakhutdinov, A. Torralba, J. Tenenbaum. Learning to Share Visual Appearance for Multiclass Object Detection. In CVPR, 2011.