Geodesic Flow Kernel for Unsupervised Domain Adaptation

Boqing Gong

University of Southern California

Joint work with Yuan Shi, Fei Sha, and Kristen Grauman

Motivation

Mismatch between different domains/datasets

- Object recognition
  - *Ex. [Torralba & Efros'11, Perronnin et al.'10]*
- Video analysis
  - *Ex. [Duan et al.'09, 10]*
- Pedestrian detection
  - *Ex. [Dollár et al.'09]*
- Other vision tasks

Performance degrades significantly!

Images from [Saenko et al.'10].

Unsupervised domain adaptation

- Source domain (labeled)

$$\mathrm{D}_S = \{(x_i, y_i), i = 1, 2, \cdots, N\} \sim P_S(X, Y)$$

- Target domain (unlabeled)

$$\mathrm{D}_T = \{(x_i, \textbf{?}), i = 1, 2, \cdots, M\} \sim P_T(X, Y)$$

*The two distributions are **not** the same!*

- **Objective**

Train classification model to work well on the target

Challenges

- How to optimally, w.r.t. target domain,

  define discriminative loss function

  select model, tune parameters

- How to solve this ill-posed problem?

  impose additional structure

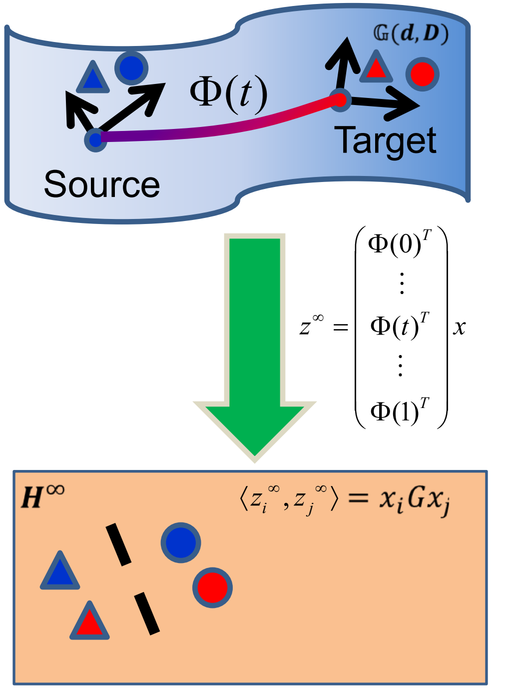
Examples of existing approaches

- Correcting sample bias
  - Ex. *[Shimodaira'00, Huang et al.'06, Bickel et al.'07]*
  - Assumption: marginal distributions are the only difference.
- Learning transductively
  - Ex. *[Bergamo & Torresani'10, Bruzzone & Marconcini'10]*
  - Assumption: classifiers have high-confidence predictions across domains.
- Learning a shared representation
  - Ex. *[Daumé III'07, Pan et al.'09, Gopalan et al.'11]*
  - Assumption: a latent feature space exists in which classification hypotheses fit both domains.

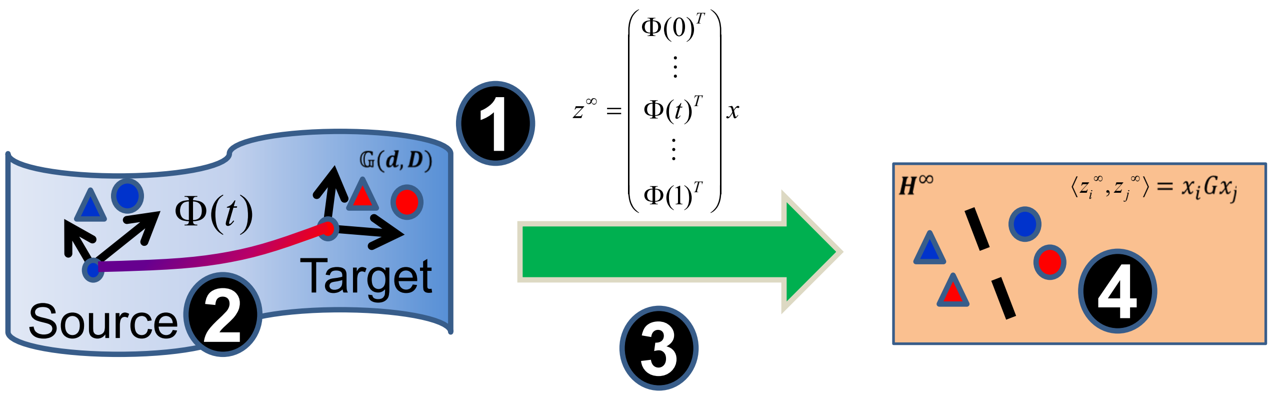
Our approach: learning a shared representation

Key insight: bridging the gap

- Fantasize infinite number of domains
- Integrate out analytically idiosyncrasies in domains
- Learn invariant features by constructing kernel



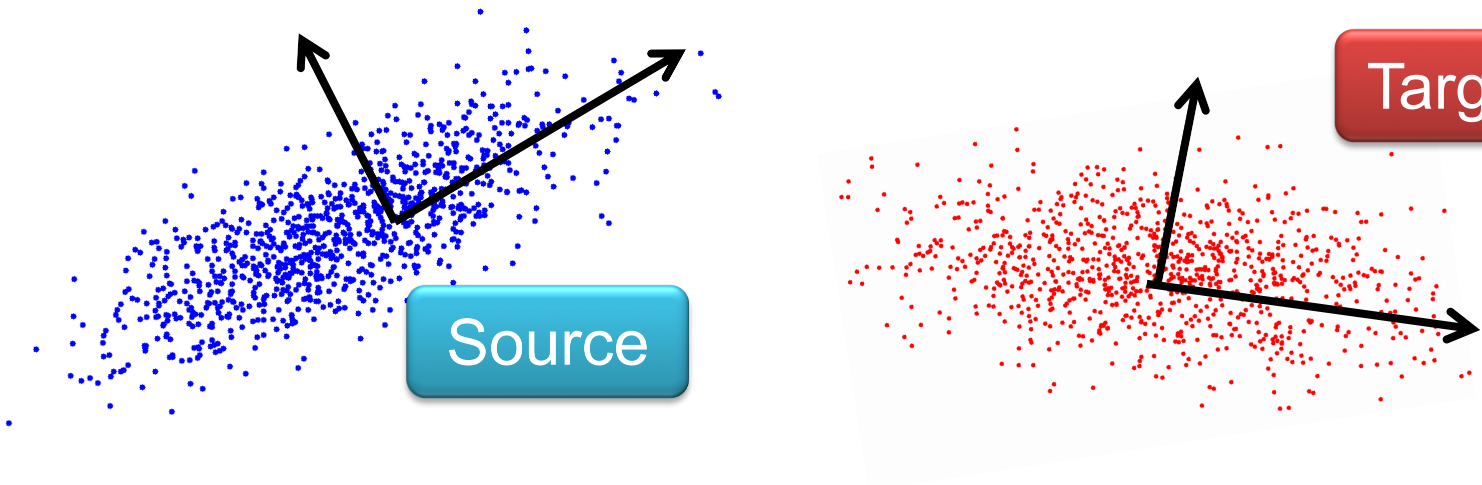
Main idea: geodesic flow kernel



1. Model data with linear subspaces
2. Model domain shift with geodesic flow
3. Derive domain-invariant features with kernel
4. Classify target data with the new features

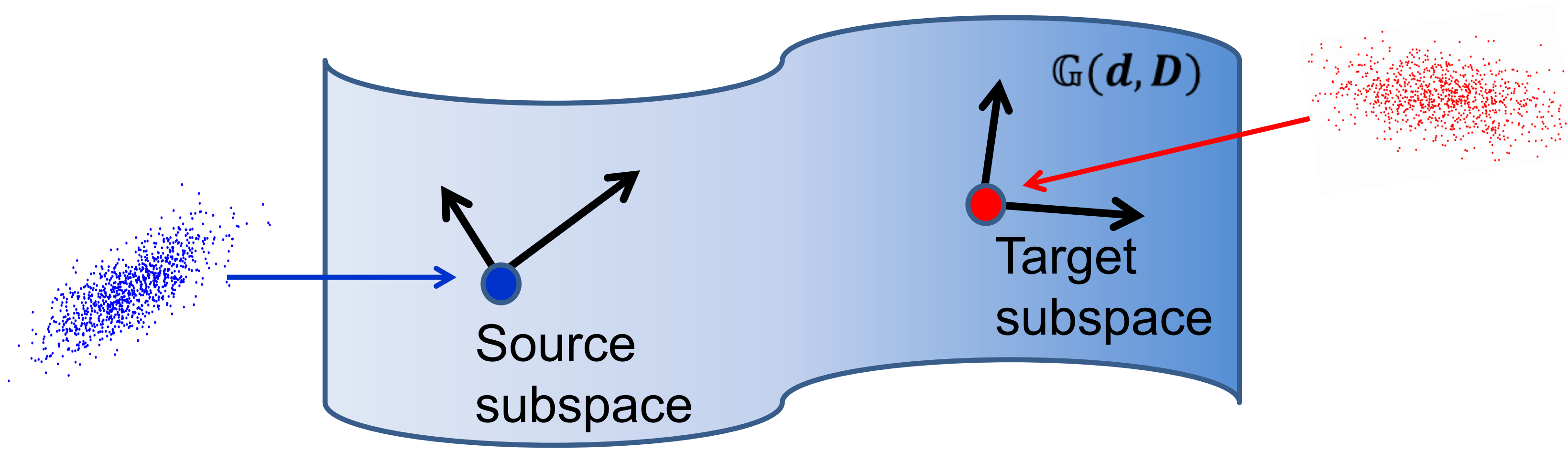
Modeling data with linear subspaces

Assume low-dimensional structure



Ex. PCA, Partial Least Squares (source only)

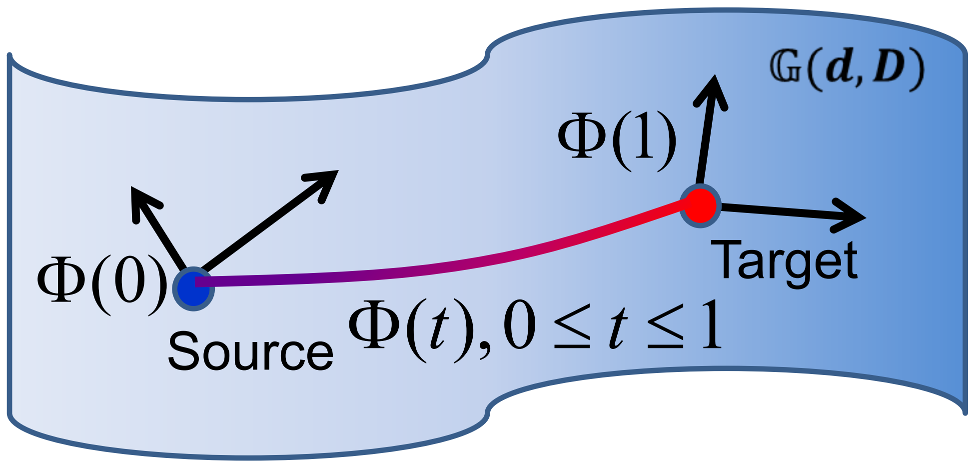
Characterizing domains geometrically



Grassmann manifold $\mathrm{G}(d, D)$

– Collection of *d*-dimensional subspaces of a vector space $\mathbf{R}^D$ $(d < D)$
– Each point corresponds to a subspace

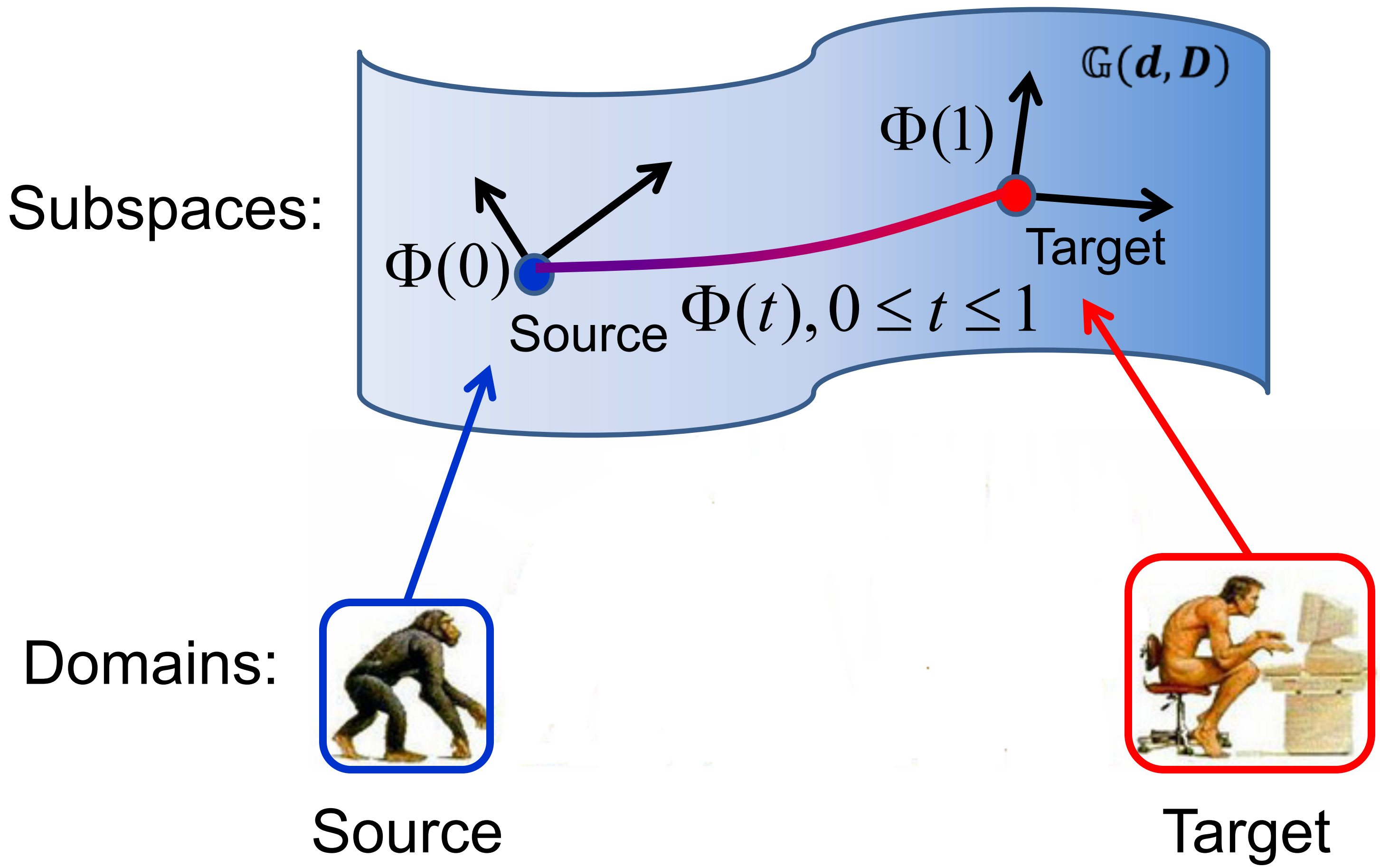
Modeling domain shift with geodesic flow



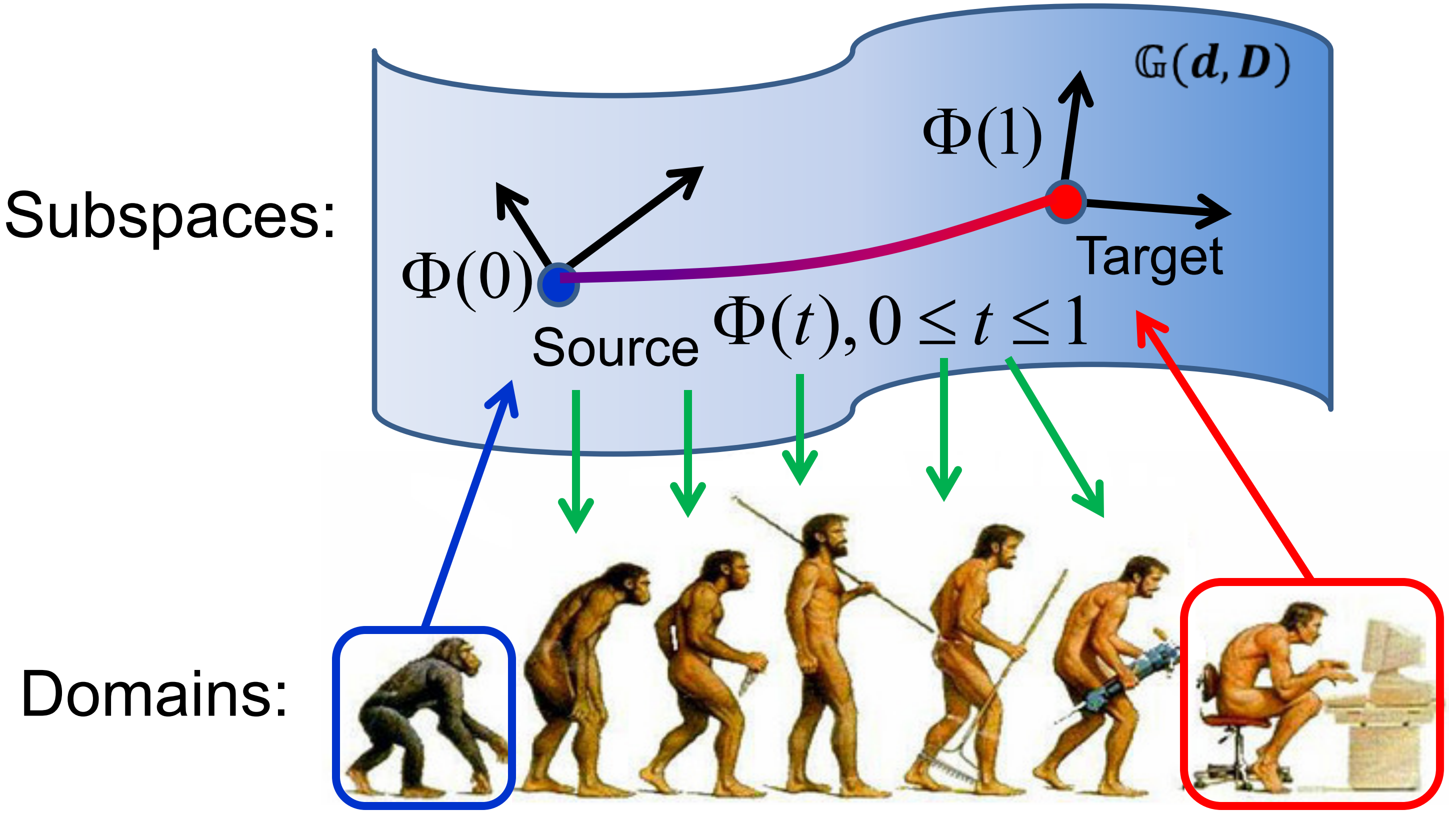
Geodesic flow on the manifold

- starting at source & arriving at target in unit time
- flow parameterized with one parameter
- closed-form, easy to compute with SVD

Modeling domain shift with geodesic flow



Modeling domain shift with geodesic flow

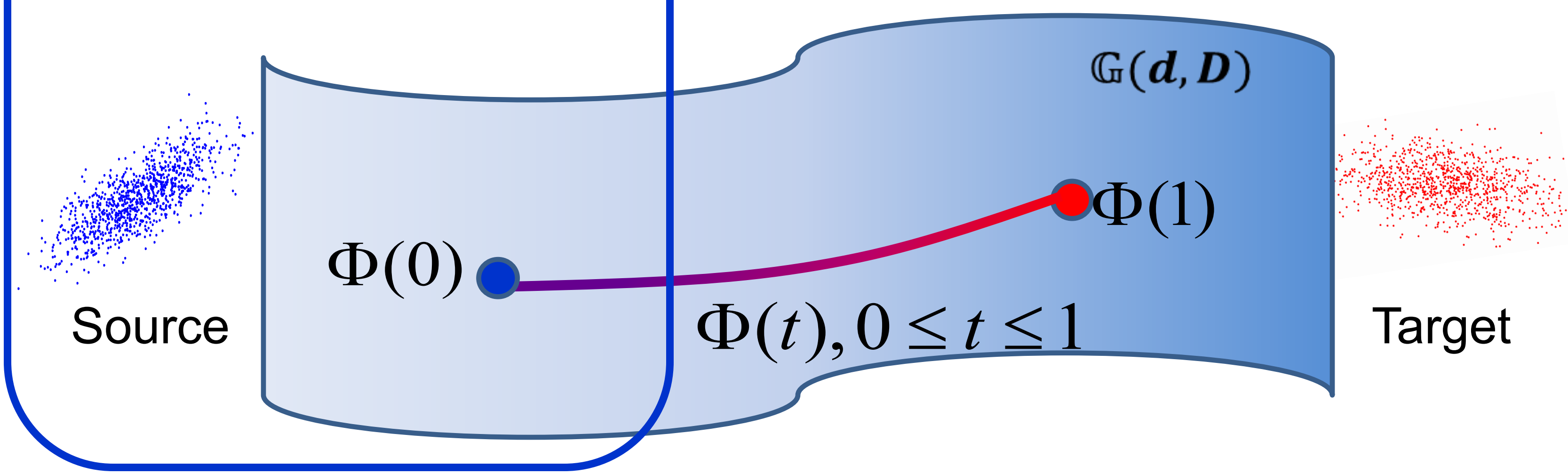


Along this flow,

points (subspaces) represent intermediate domains.

Domain-invariant features

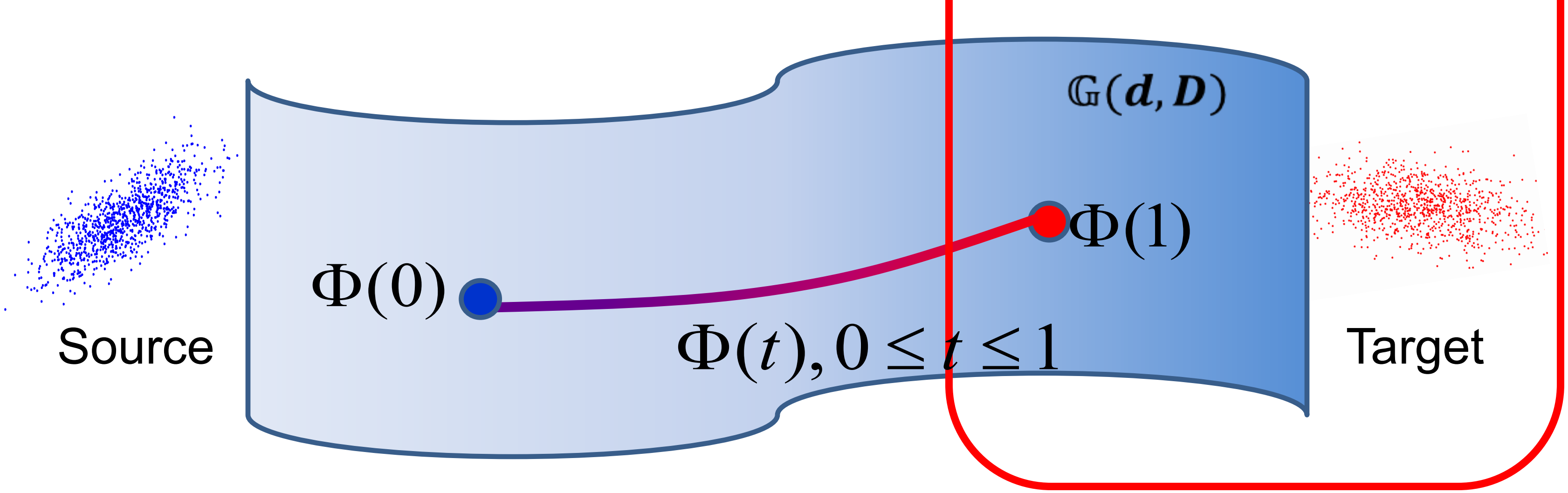
$$z^{\infty} = [\Phi(0)^T x, \cdots, \Phi(t)^T x, \cdots, \Phi(1)^T x]$$



More similar to source.

Domain-invariant features

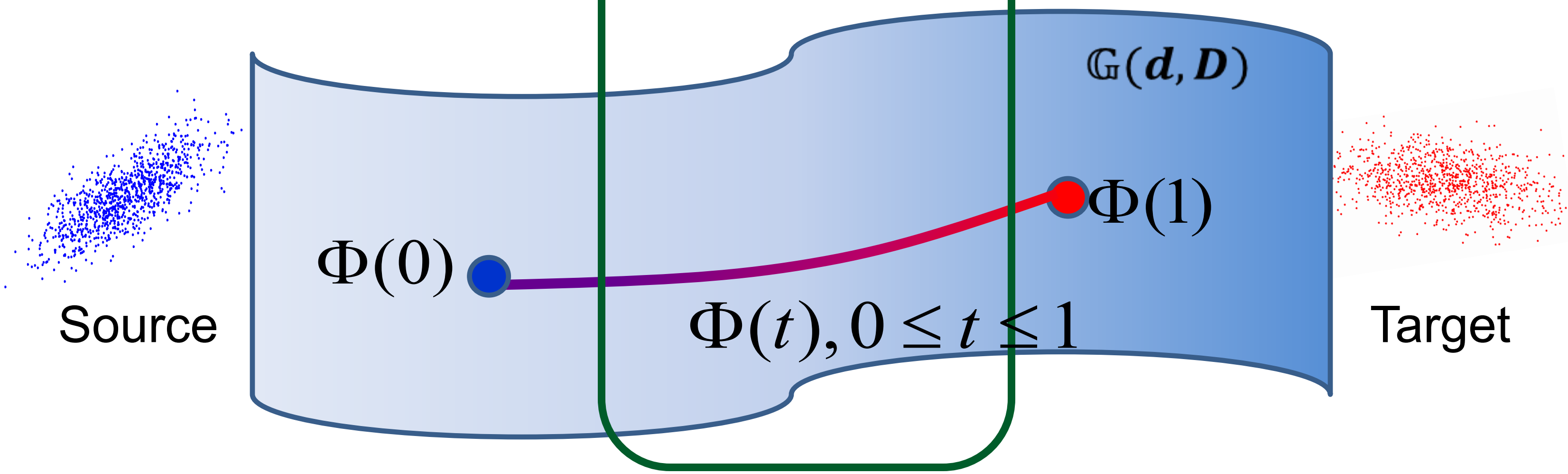
$$z^{\infty} = [\Phi(0)^T x, \cdots, \Phi(t)^T x, \cdots, \Phi(1)^T x]$$



More similar to target.

Domain-invariant features

$$z^{\infty} = [\Phi(0)^T x, \cdots, \Phi(t)^T x, \cdots, \Phi(1)^T x]$$



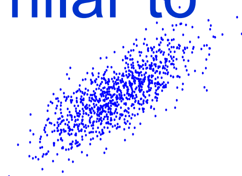
Blend the two.

Measuring feature similarities with inner products

$$z_i^\infty = [\Phi(0)^T x_i, \cdots, \Phi(t)^T x_i, \cdots, \Phi(1)^T x_i]$$

$$z_j^\infty = [\Phi(0)^T x_j, \cdots, \Phi(t)^T x_j, \cdots, \Phi(1)^T x_j]$$

More similar to source.



More similar to target.

$\langle z_i^\infty, z_j^\infty \rangle$: Invariant to either source or target.

Learning domain-invariant features with kernels

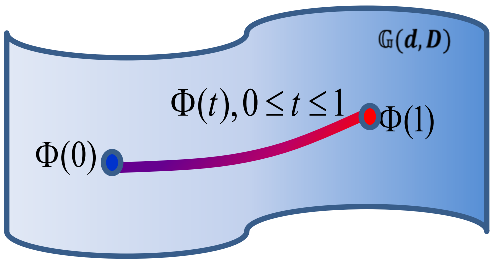
We define the **geodesic flow kernel (GFK)**:

$$\langle z_i^{\infty}, z_j^{\infty} \rangle = \int_0^1 (\Phi(t)^T x_i)^T (\Phi(t)^T x_j) dt = x_i^T G x_j$$

- **Advantages**
  - Analytically computable
  - Robust to variants towards either source or target
  - Broadly applicable: can kernelize many classifiers

Contrast to discretely sampling

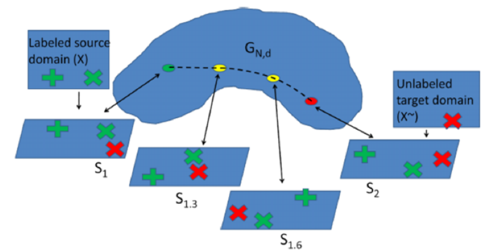
GFK (ours)



$$\langle z_i^{\infty}, z_j^{\infty} \rangle = \int_0^1 (\Phi(t)^T x_i)^T (\Phi(t)^T x_j) dt = x_i^T G x_j$$

No free parameters

[Gopalan et al. ICCV 2011]

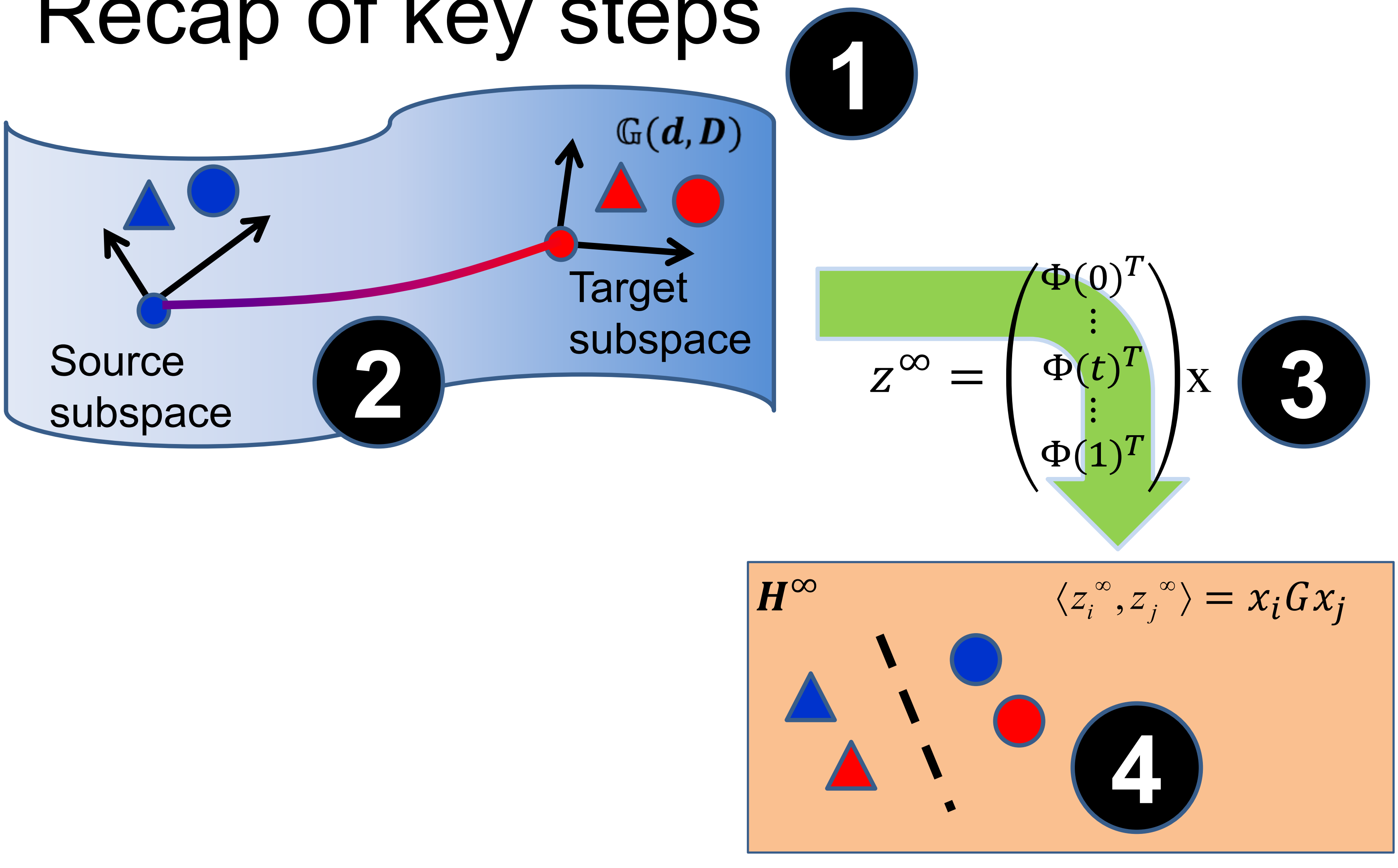


Dimensionality reduction

Number of subspaces, dimensionality of subspace, dimensionality after reduction

GFK is conceptually cleaner and computationally more tractable.

Recap of key steps



Experimental setup

- Four domains
- Features
  Bag-of-SURF
- Classifier: 1NN
- Average over 20 random trials

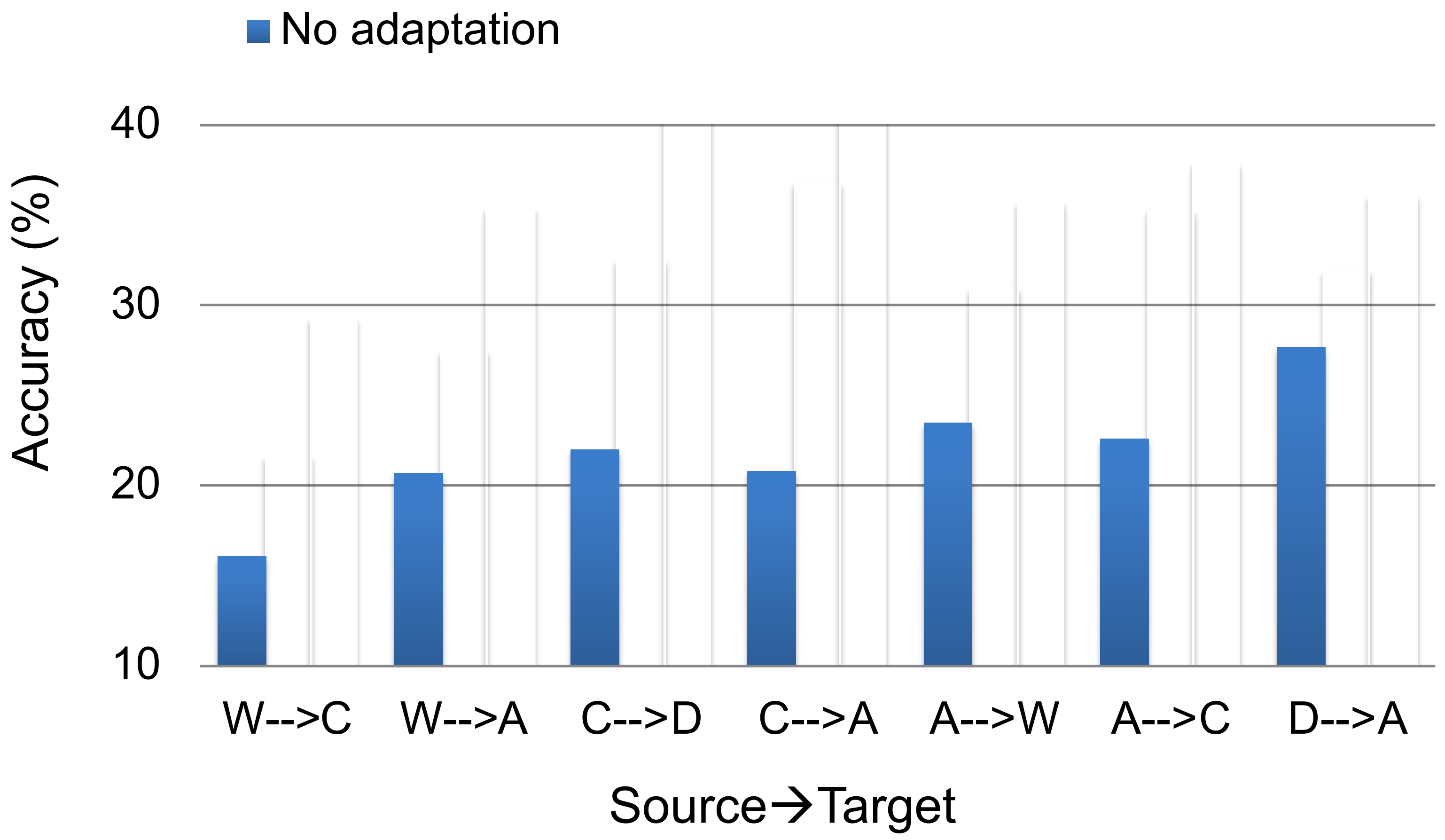
Caltech-256

Amazon

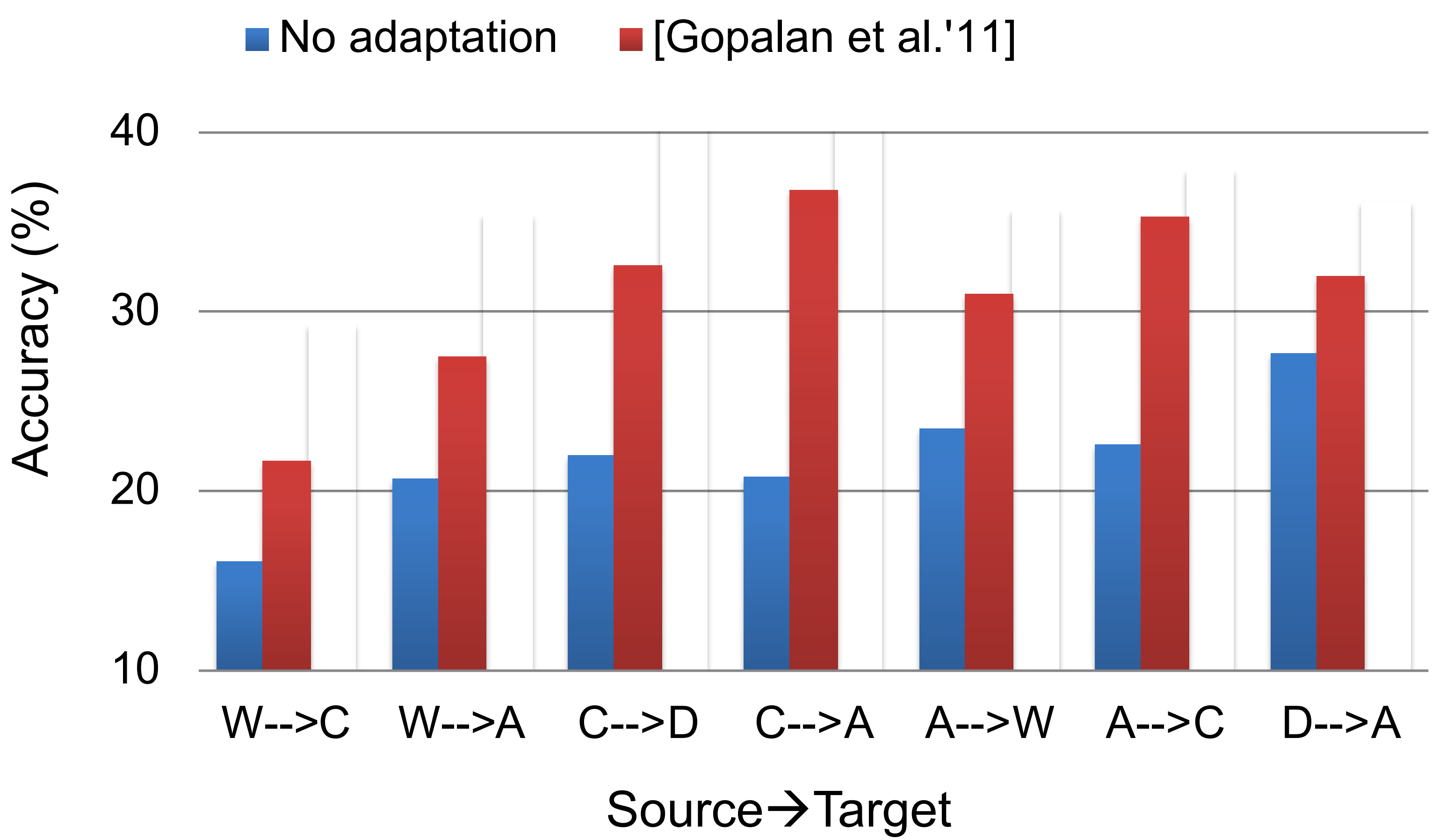
DSLR

Webcam

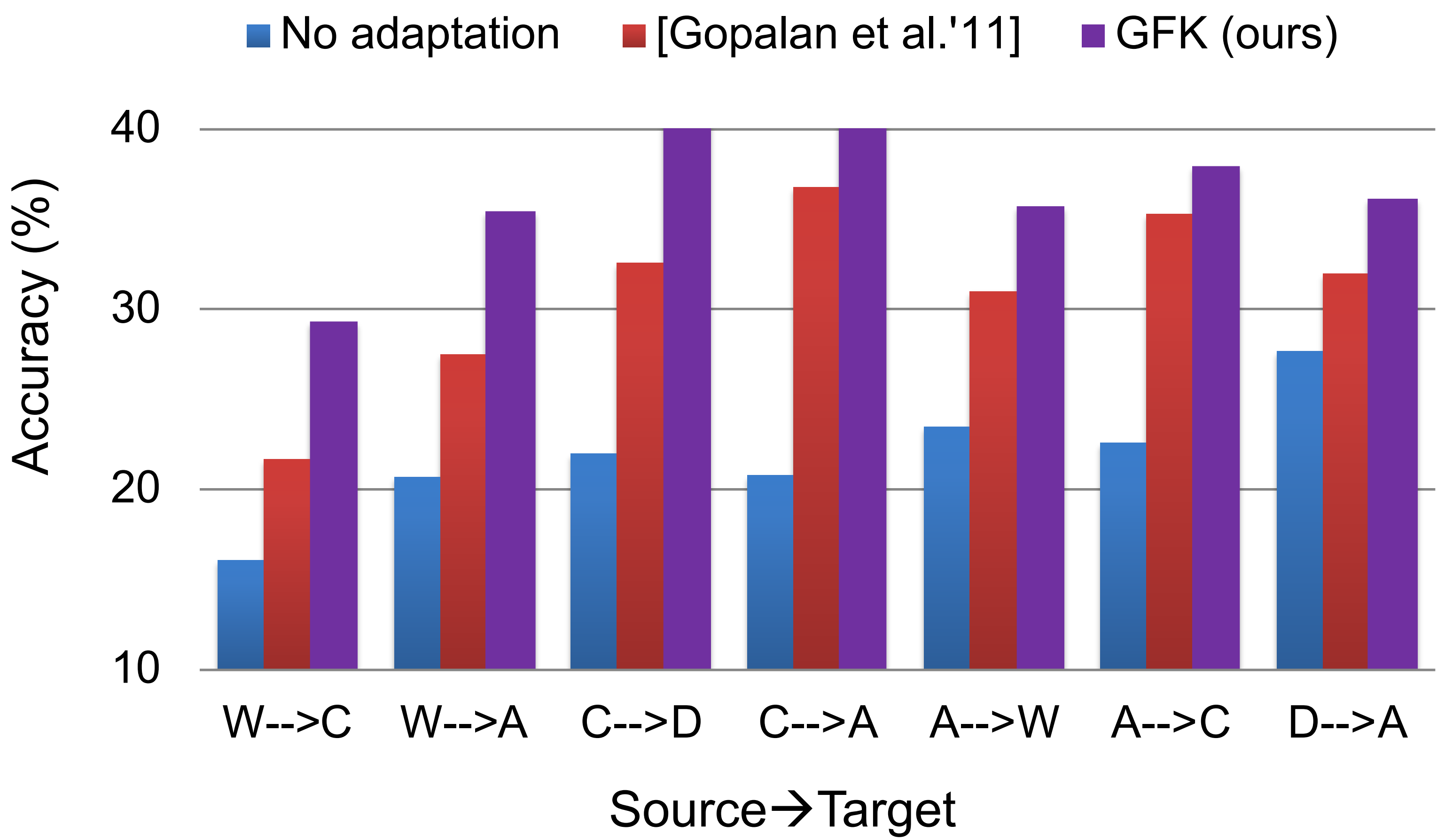
Classification accuracy on target



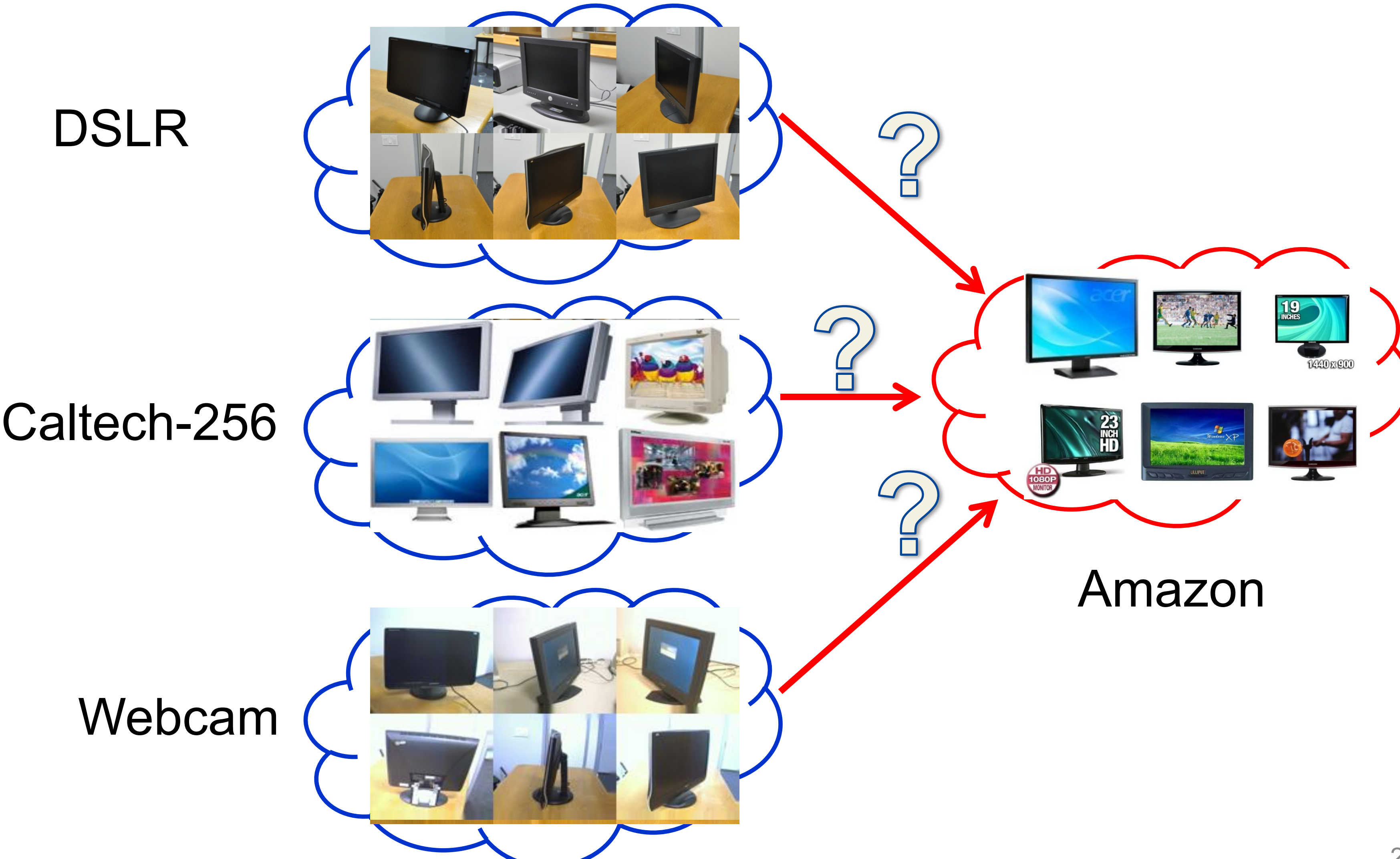
Classification accuracy on target



Classification accuracy on target



Which domain should be used as the source?



Automatically selecting the best

We introduce the **Rank of Domains** measure:

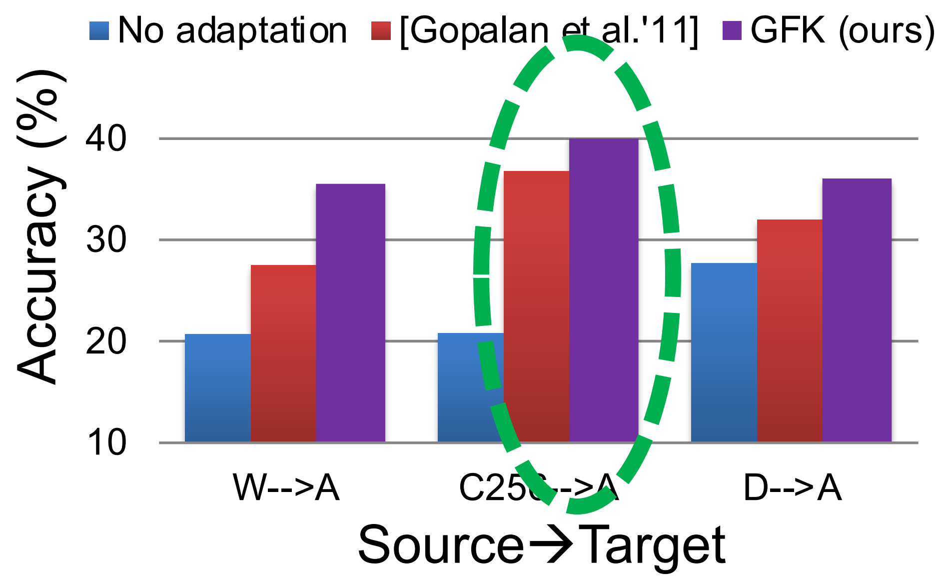
$$\mathcal{R}(\mathcal{S},\mathcal{T}) = \frac{1}{\mathrm{d}^*}\sum_{i}^{\mathrm{d}^*}\theta_i \left[KL(\mathcal{S}_i\|\mathcal{T}_i) + KL(\mathcal{T}_i\|\mathcal{S}_i)\right]$$

Intuition

- Geometrically, how subspaces disagree
- Statistically, how distributions disagree

Automatically selecting the best

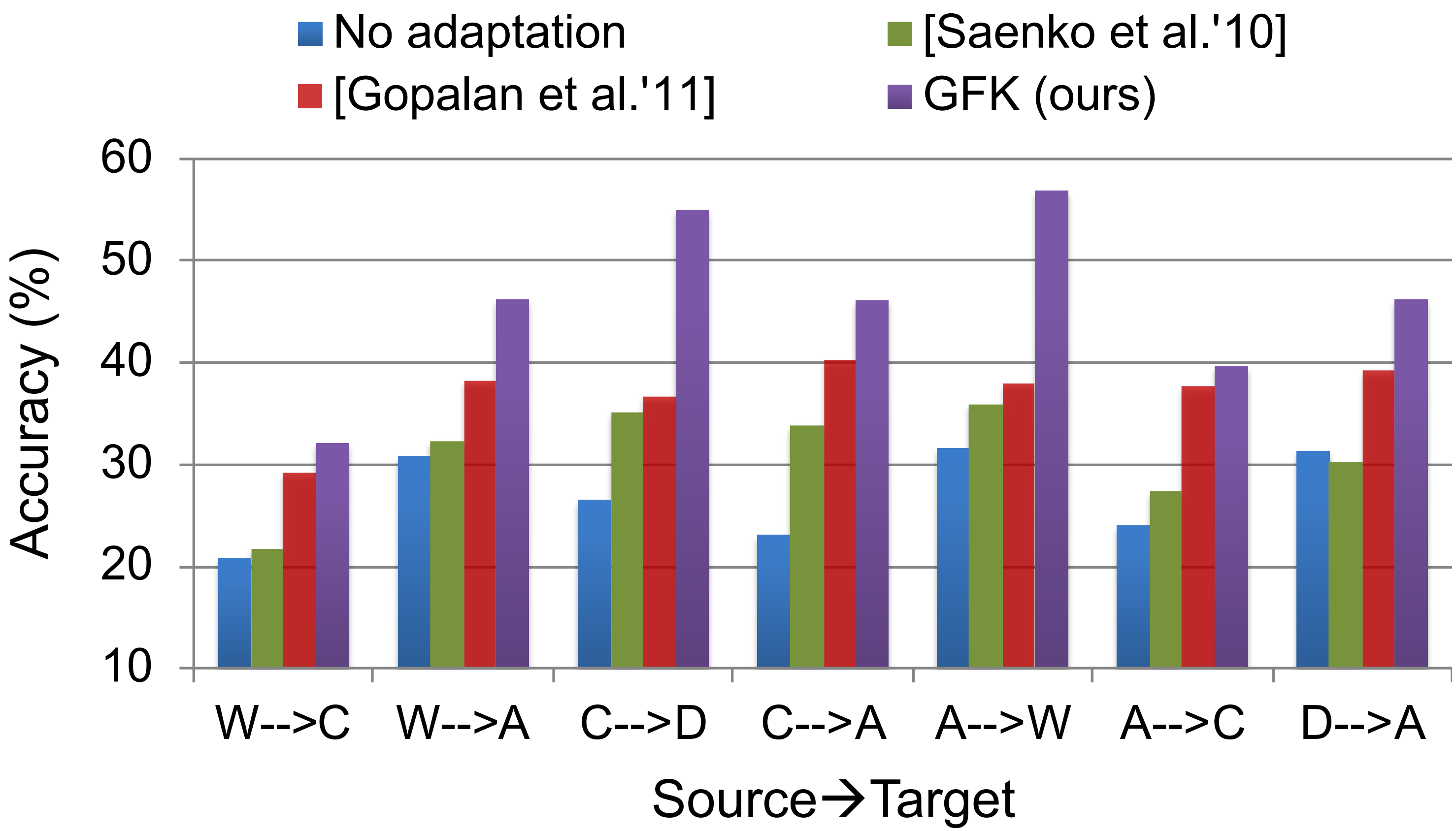
Possible sources	Our ROD measure
Caltech-256	0.003
Amazon	0
DSLR	0.26
Webcam	0.05



Caltech-256 adapts the best to Amazon.

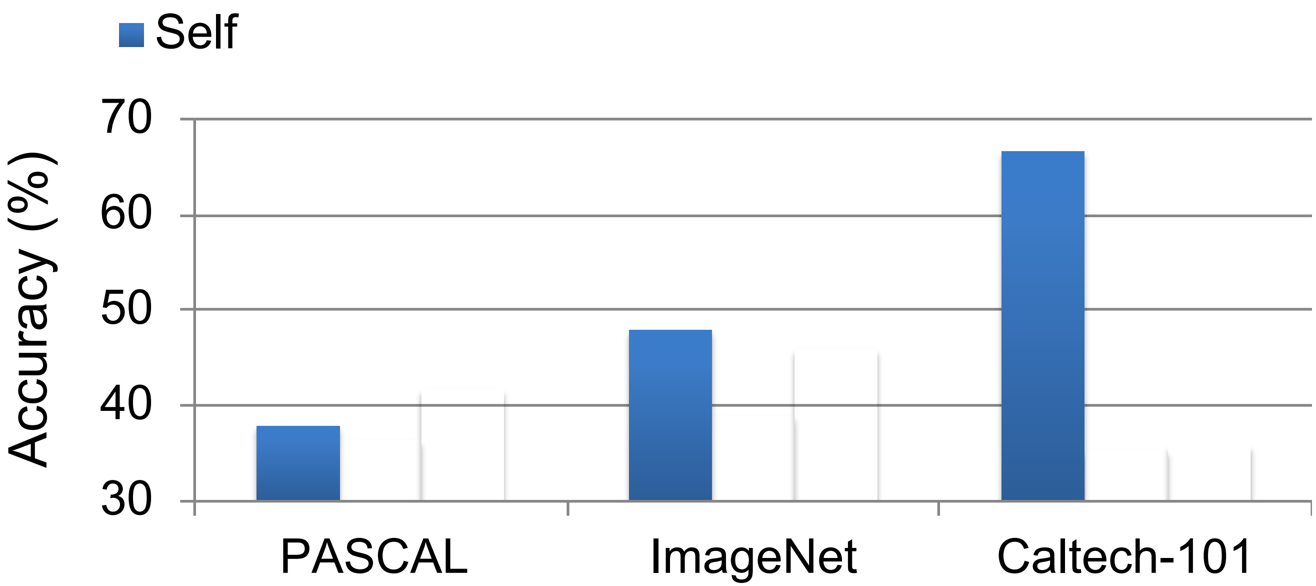
Semi-supervised domain adaptation

Label three instances per category in the target



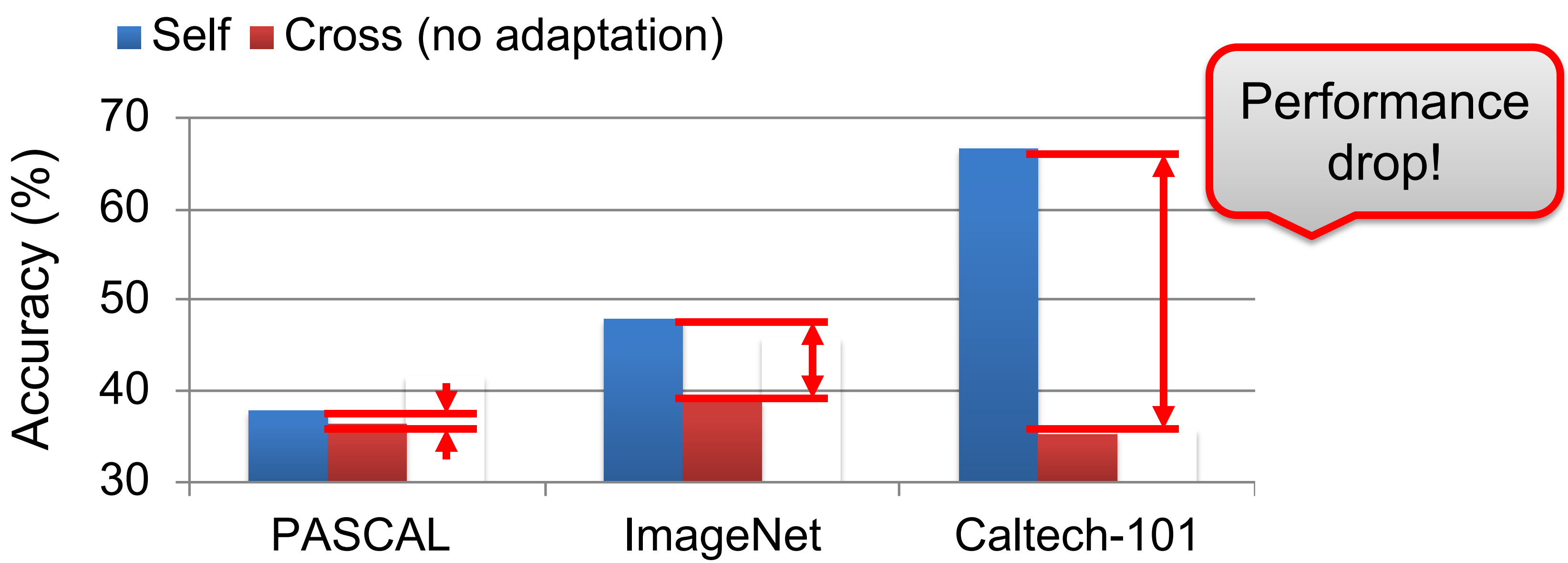
Analyzing datasets in light of domain adaptation

Cross-dataset generalization *[Torralba & Efros'11]*



Analyzing datasets in light of domain adaptation

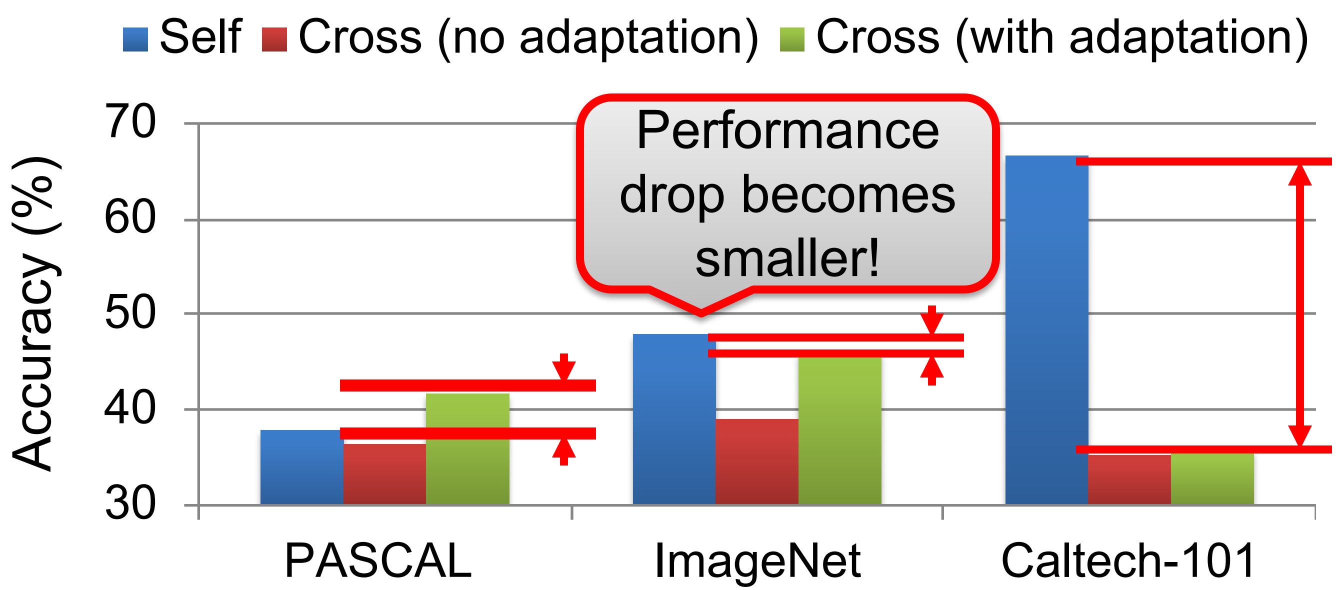
Cross-dataset generalization *[Torralba & Efros'11]*



Caltech-101 generalizes the worst.
Performance drop of ImageNet is big.

Analyzing datasets in light of domain adaptation

Cross-dataset generalization *[Torralba & Efros'11]*



Caltech-101 generalizes the worst (w/ or w/o adaptation).
There is nearly no performance drop of ImageNet.

Summary

- Unsupervised domain adaptation
  - Important in visual recognition
  - Challenge: no labeled data from the target
- Geodesic flow kernel (GFK)
  - **Conceptually clean formulation**: no free parameter
  - **Computationally tractable**: closed-form solution
  - **Empirically successful**: state-of-the-art results
- New insight on vision datasets
  - Cross-dataset generalization with domain adaptation
  - Leveraging existing datasets despite their idiosyncrasies

Future work

- Beyond subspaces

  Other techniques to model domain shift

- From GFK to statistical flow kernel

  Add more statistical properties to the flow

- Applications of GFK

  Ex., face recognition, video analysis

Summary

Thank you!

- Unsupervised domain adaptation
  - Important in visual recognition
  - Challenge: no labeled data from the target
- Geodesic flow kernel (GFK)
  - Conceptually clean formulation
  - Computationally tractable
  - Empirically successful
- New insight on vision datasets
  - Cross-dataset generalization with domain adaptation
  - Leveraging existing datasets despite their idiosyncrasies