ACCV 2012

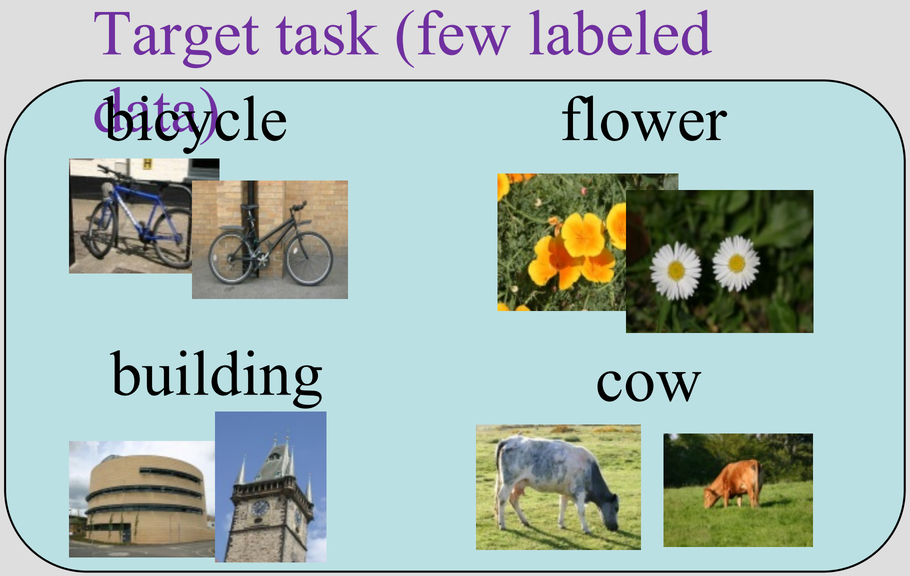
Cross-Database Transfer Learning via Learnable and Discriminant Error-Correcting Output Codes

Feng-Ju Chang[1] Yen-Yu Lin[1] Ming-Fang Weng[2]

[1]Research Center for Information Technology Innovation, Academia Sinica, Taiwan
[2]Institute of Information Science, Academia Sinica, Taiwan

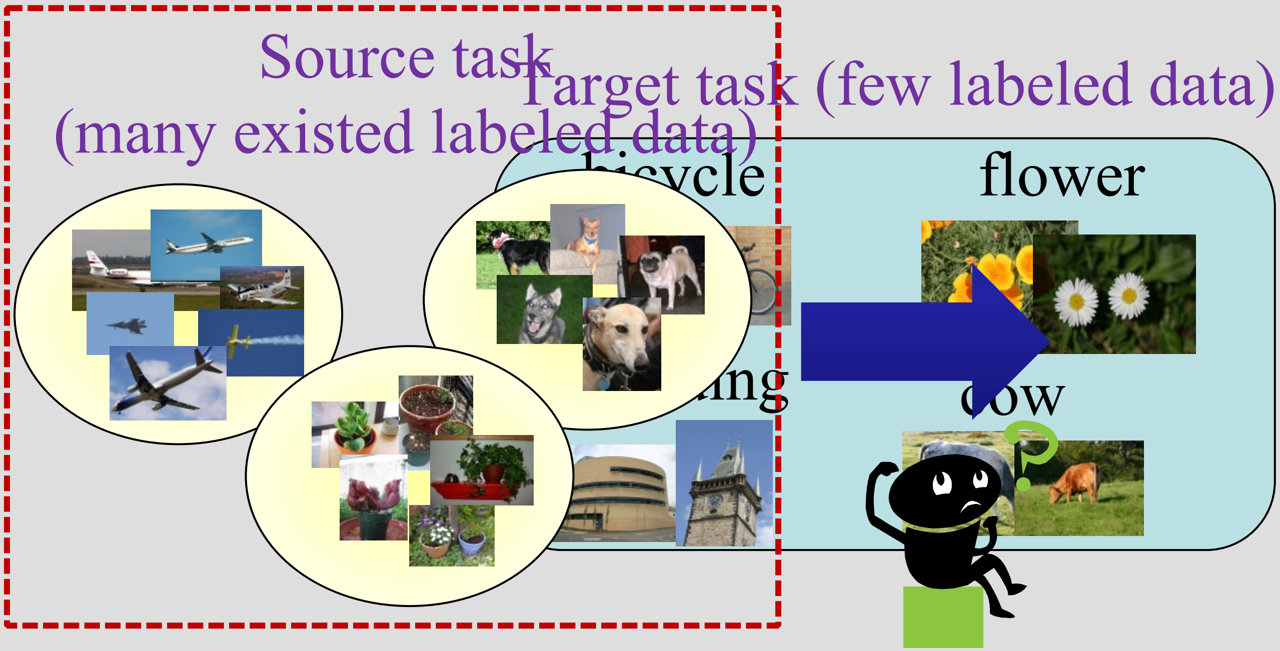
The Problem

- Multi-class object recognition with few labeled data
  - Goal: Learn a target classifier with low generalization errors



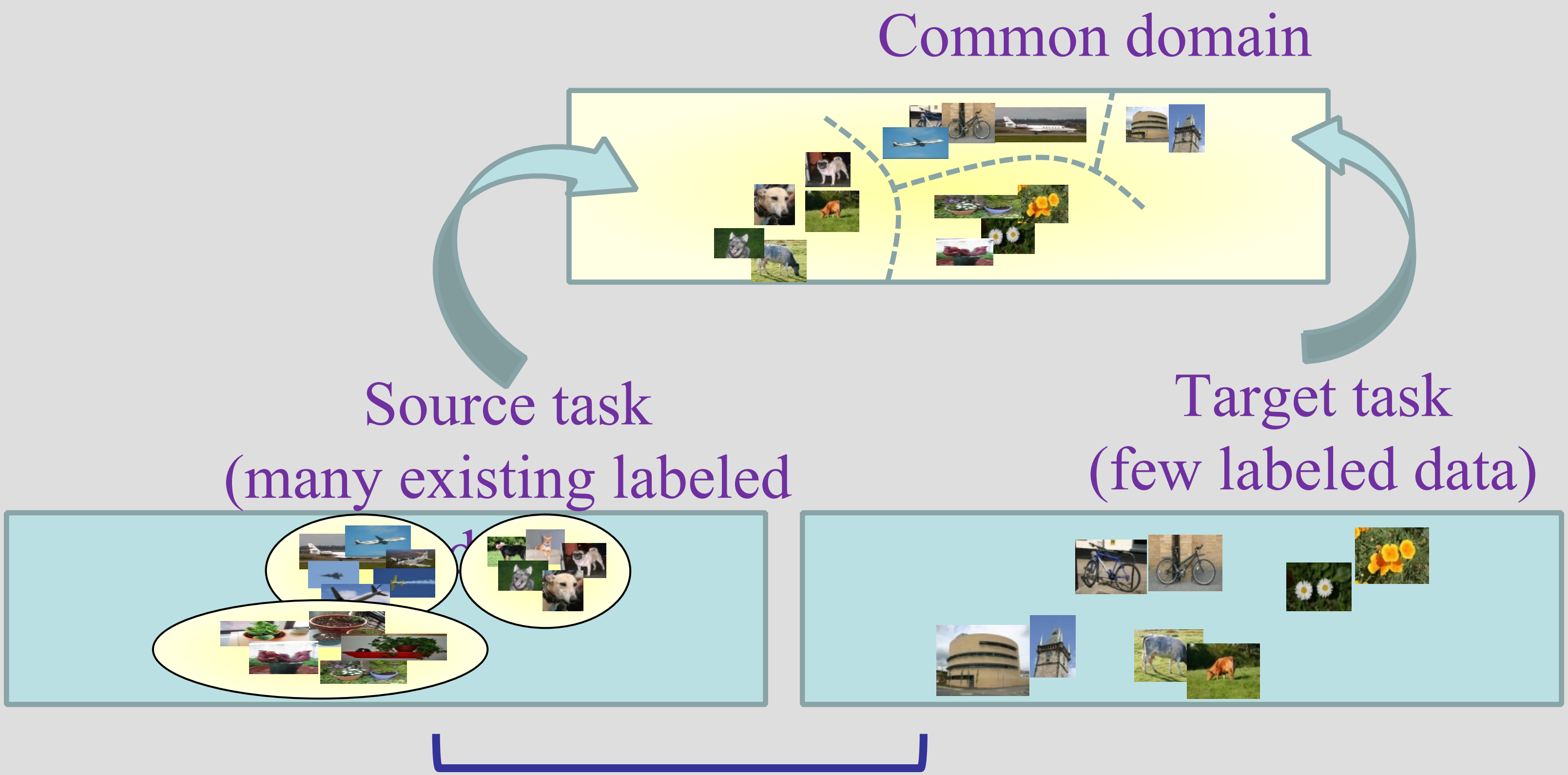
Over-fitting
(poor generalization)

- What and how prior knowledge help to learn a robust classifier without labeling new data?



Can be from different categories & different databases

Motivation

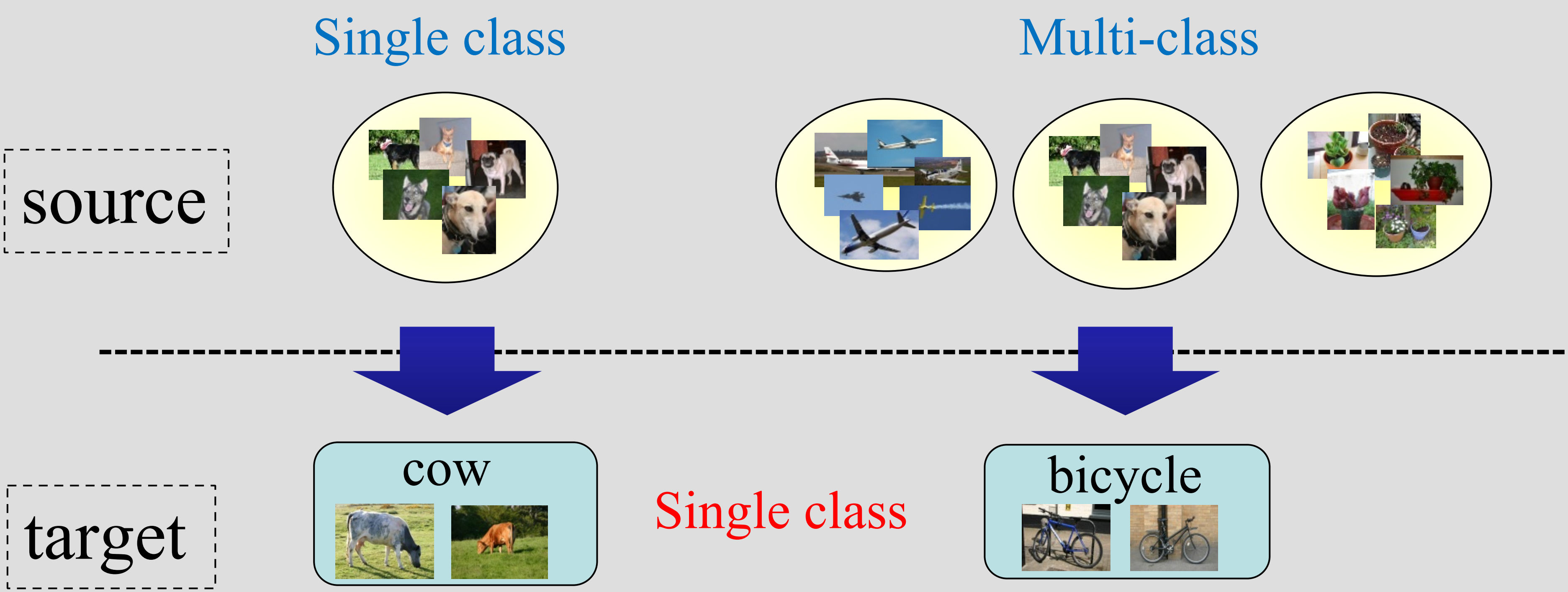


different categories, database, feature distribution

Previous work: Instance transfer, feature transfer, model transfer

Motivation

- Conventional TL algorithms: Lack multi-class formulation
  - ➢ Cannot explore intra- and inter- class distribution (less discriminant)



[Dai et al., ICML’07]

[Fei-Fei et al., TPAMI’06]

[Yan et al., MM’07]

[Yao et al., CVPR’12
]

Motivation

- Our Approach: Attribute Transfer

can be shared
by several classes

can reach
zero-shot learning

[Lampert et al., CVPR'09]

Motivation

- Our goal: Attributes should be learnable
  - Alleviate the labeling effort

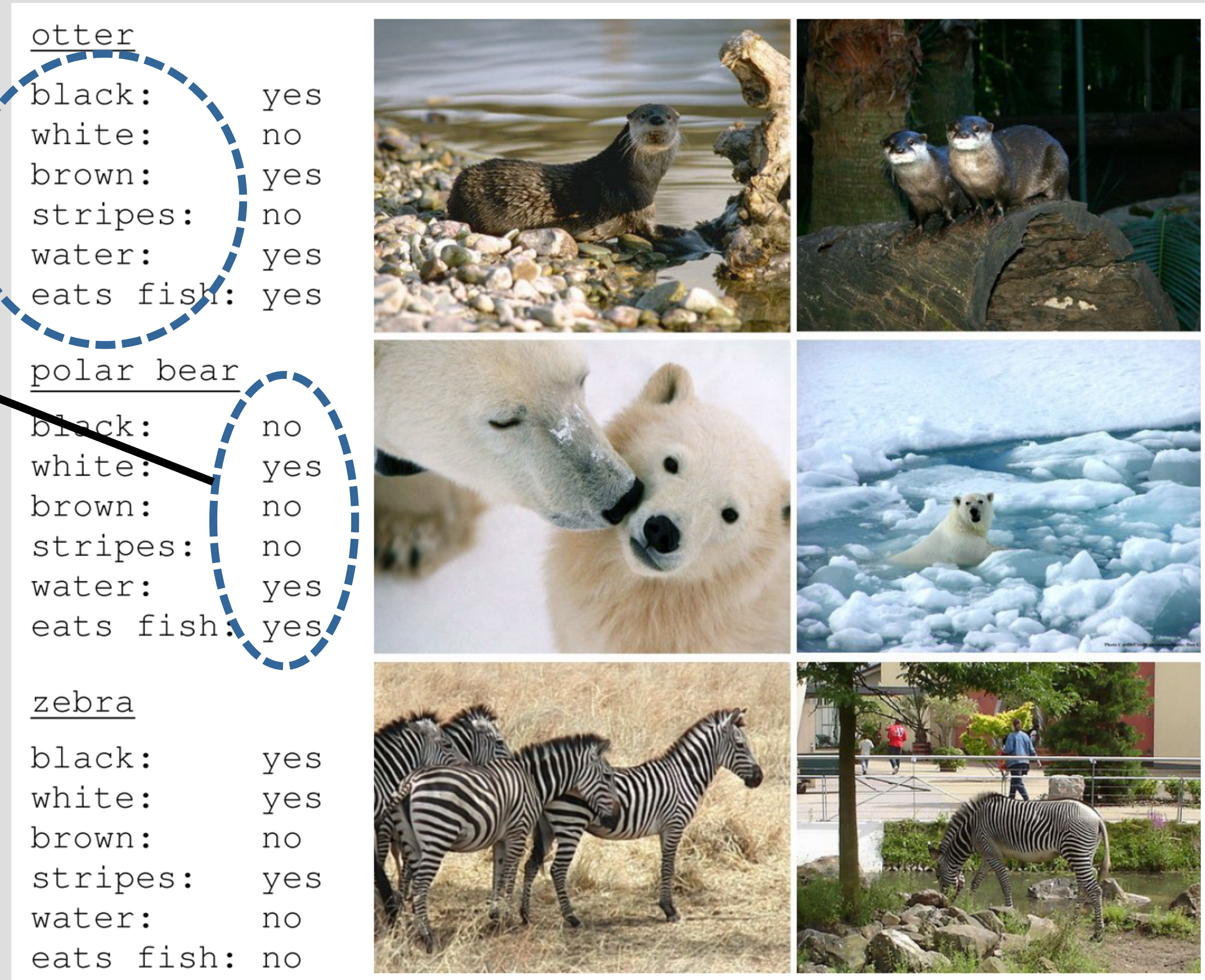
Predefined attributes

Attribute labeling

Inapplicable in different datasets!

Expensive human efforts!

[Lampert et al., CVPR'09]



Our Approach

- Multi-classes (source) to multi-classes (target) knowledge transfer

- What to transfer: a sequence of learnable, discriminant attributes
  - Commonly shared by the source and target domains
  - Converted two multi-class classification tasks to related, binary ones

- How to transfer: Two-layer multi-task variant of AdaBoost.OC
  - Boosting algorithm with error-correcting output codes (ECOC)
    - Better generalization [Dietterich et al., JAIR'95]
  - Outer layer: Attribute partition discovery
  - Inner layer: Attribute classifier learning

Characteristics of Our Approach

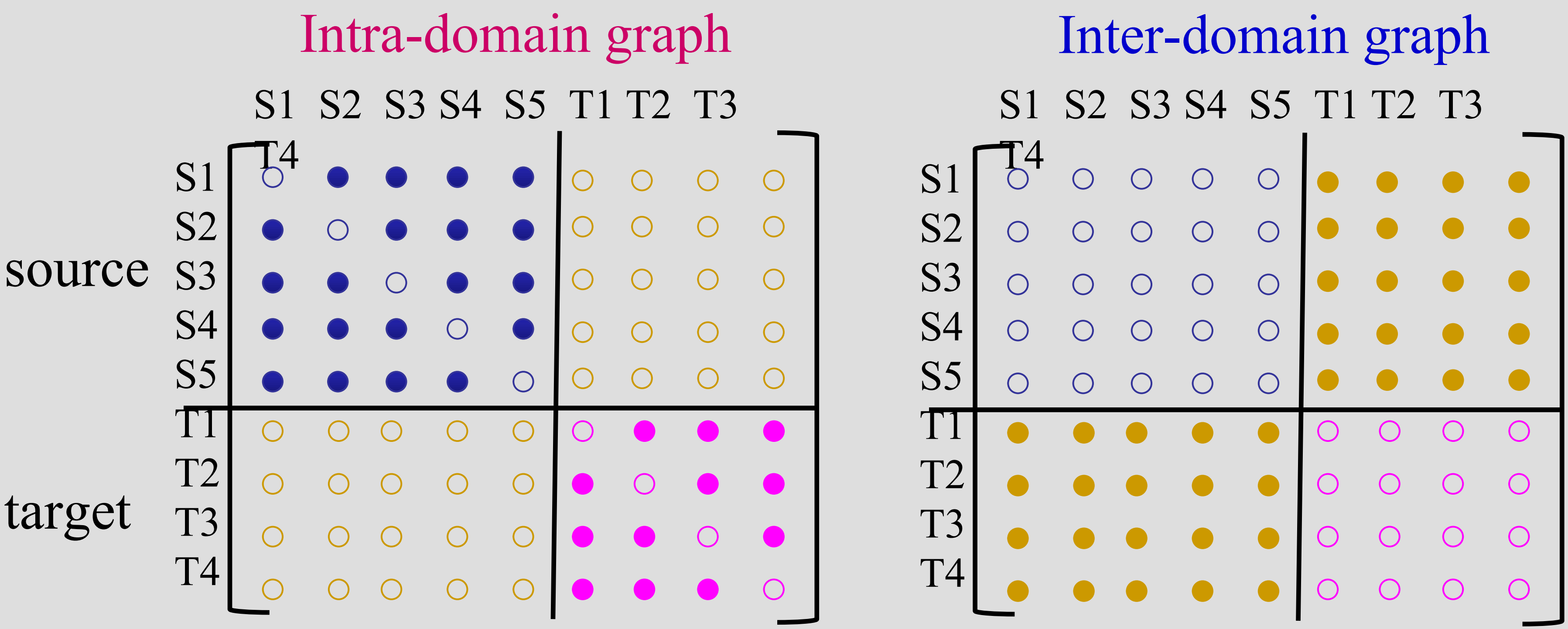
- Two-layer multi-task variant of AdaBoost.OC
  - Outer layer: Discover a set of attribute partitions
    - Discriminant: Multi-class formulation
    - Learnable: Without human effort
    - Complementary: Iterative error minimization
  - Inner layer: Learn attribute classifiers
    - Employ classifier sharing principle
    - Support multiple kernel learning: Combining various low-level features

Outline

- Multi-class transfer learning
  - Connect two domains via attribute partition sharing
  - Knowledge transfer by jointly learning classifiers
- Experimental results

Sharing Attribute Partions for Domain Correlation

- Two criteria for designing the partition function:
  - Criterion 1: Discriminant for both domains
  - Criterion 2: Achieve class consistency between the opposite domains



Establish Partition Function

- Combine above two graphs, the partition function B can be learned by the following optimization problem:

$$B^* = \arg\max_B J_{dis}(B)/J_{shr}(B), \text{ where}$$

$J_{dis}(B) =$ Criterion 1: Maximize discriminability of intra domains

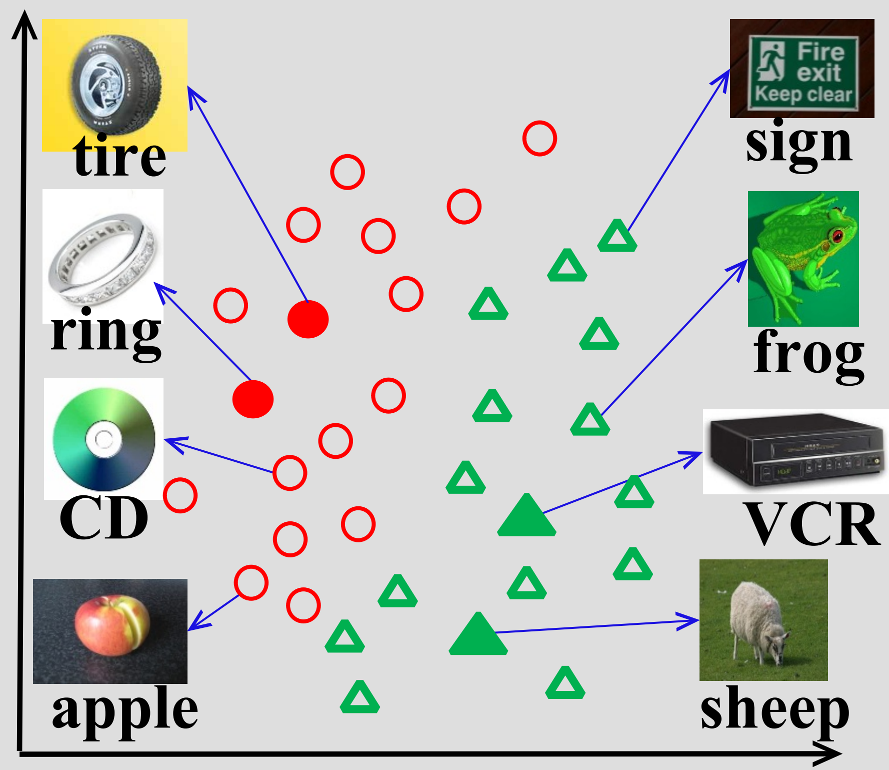
$J_{shr}(B) =$ Criterion 2: Minimize the class inconsistency between two domains

  - Binary integer programming: NP-complete problem
  - Continuous relaxation: Solving a generalized eigenvalue problem

How to Transfer

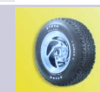
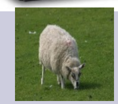
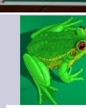
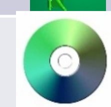
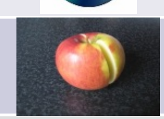
- Two-layer multi-task variant of AdaBoost.OC
  - Outer layer: Discover a set of attribute partitions

Attribute “circle”



Target data

Source data

Attribute Partition	
Class	circle
	1
	1
	0
	0
	0
	0
	1
	1

Outline

- Multi-class Transfer Learning
  - Connect two domains via attribute partition function sharing
  - Knowledge transfer by jointly learning classifiers
- Experimental results

Jointly Learning Classifiers for Knowledge Transfer

- Attribute classifier learning:
  - Exploit the classifier sharing principle [Torralba et al., TPAMI'07]
- A boosting-based approach is presented
  - The design of weak learners: *dyadic hypercuts*
    - The discriminative power of each kernel → a set of weak learners
    - Achieve multiple kernel learning
  - Dual-domain boosting for knowledge transfer
    - The relatedness between tasks: Modeled by the shared weak learners
    - The difference between tasks: Reflected by respective ensemble coefficients
    - Avoid negative transfer

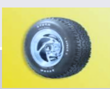
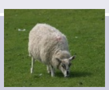
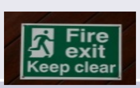
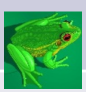
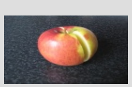
$$f^{\ell}(\mathbf{x}) = \sum_{t=1}^{V} \beta_t^{\ell} h_t(\mathbf{x}) \quad \text{for } \ell \in \{S, T\}$$

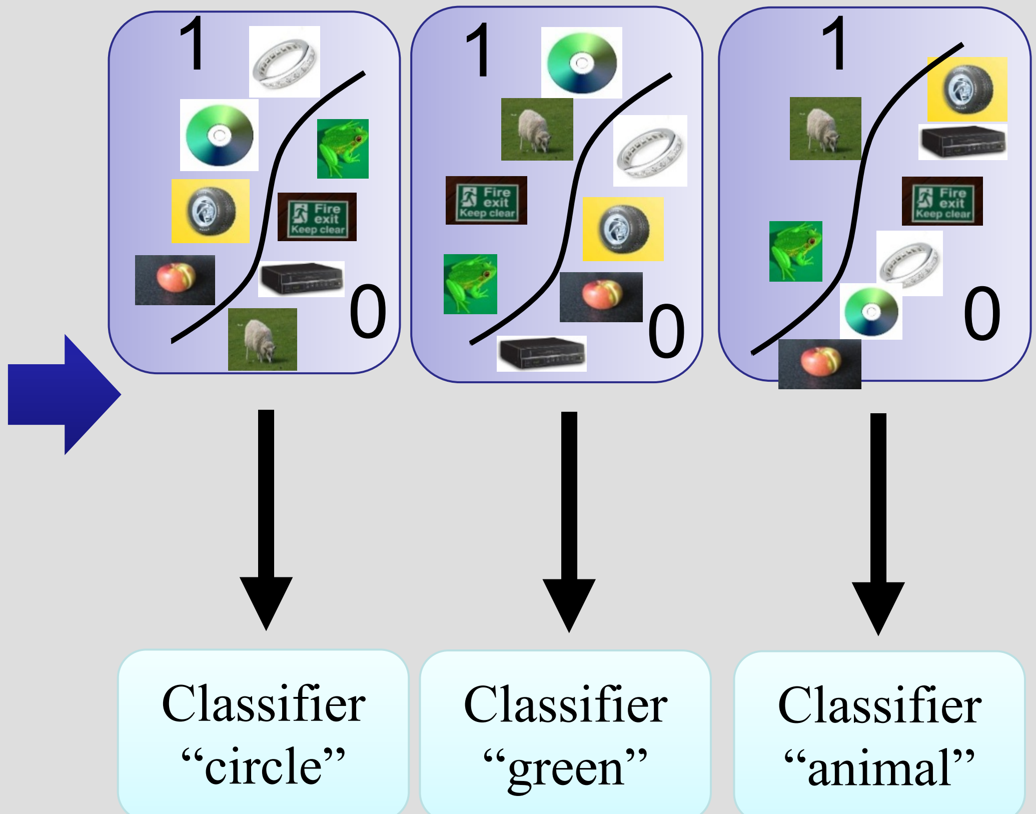
Difference

Relatedness

How to Transfer

- Two-layer multi-task variant of AdaBoost.OC
  - Outer layer: Discover a set of attribute partitions
  - Inner layer: Learn attribute classifiers

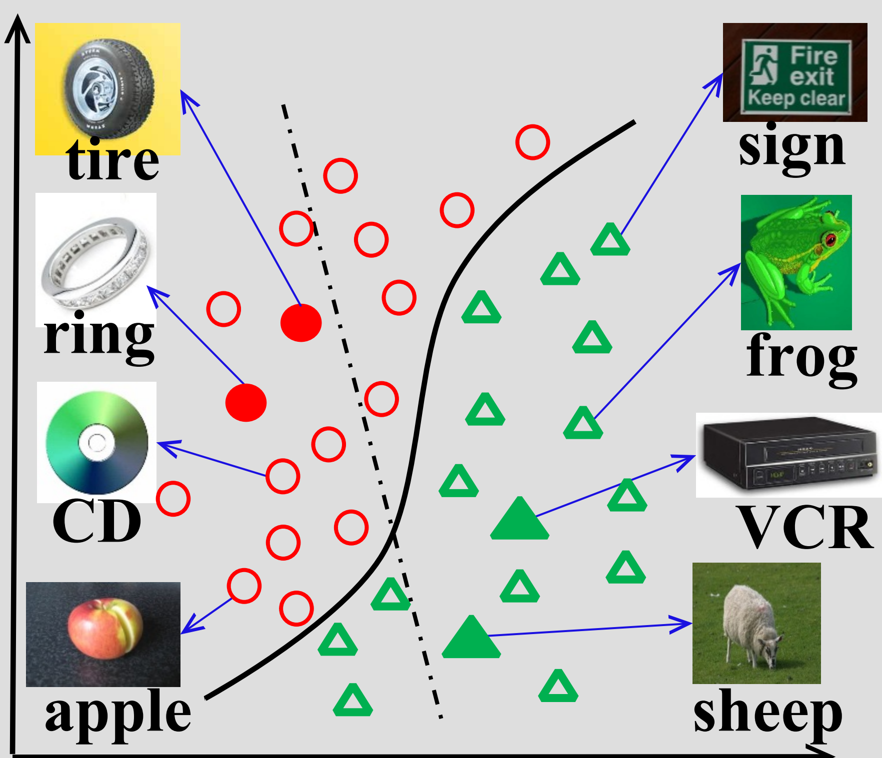
Class	circle	green	animal
	1	0	0
	1	0	0
	0	0	0
	0	1	1
	0	1	0
	0	1	1
	1	1	0
	1	0	0



How to Transfer

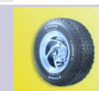
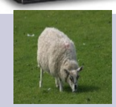
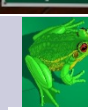
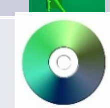
- Two-layer multi-task variant of AdaBoost.OC
  - Outer layer: Discover a set of attribute partitions
  - Inner layer: Learn attribute classifiers

Attribute "circle"



Target data

Source data

Attribute Partition	
Class	circle
	1
	1
	0
	0
	0
	0
	1
	1

Outline

- Multi-class Transfer Learning
  - Connect two domains via attribute partition function sharing
  - Knowledge transfer by jointly learning classifiers
- Experimental results

Experimental Settings

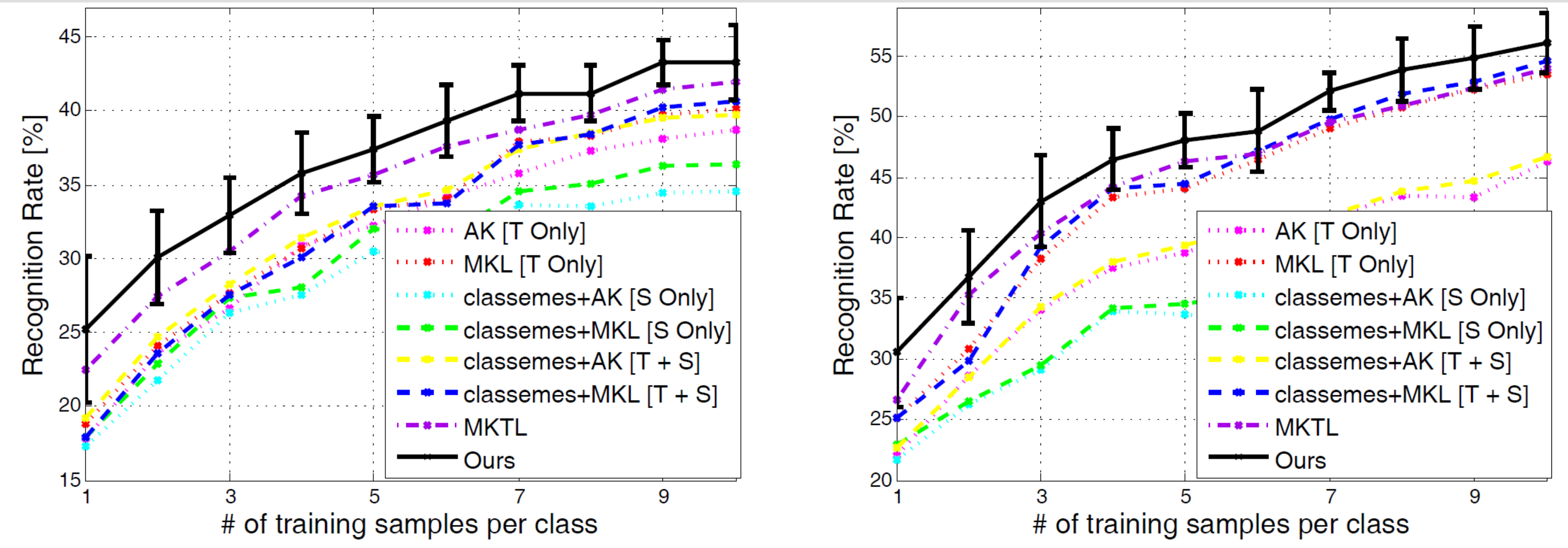
- Datasets:
  - Caltech256, SUN09, MSRC
- Features and kernels:
  - GIST, BoW-SIFT, Color histogram, Texton
  - RBF kernels with the Euclidean distance
- Baselines:
  - Target Only (T Only) :
    - Average Kernel SVM (AK), Multiple Kernel SVM (MKL)
  - Source only (S Only) :
    - Classemes [Torresani et al., ECCV'10] + AK, Classemes + MKL
  - Target + Source (T + S):
    - Classemes + AK, Classemes + MKL, Multiple Kernel Transfer Learning (MKTL) [Jie et al., ICCV'11]

Experimental Settings

- For each dataset:
  - 10 target categories and 20 source categories are randomly selected
  - Training: 1 to 10 samples per target class, 50 samples per source class
  - Testing: 50 samples per target class
- All experiments are repeated 10 times
- Two scenarios:
  - Within-database transfer learning:
    - Source and target classes are from the same database
  - Cross-database transfer learning:
    - Source and target classes are from different databases

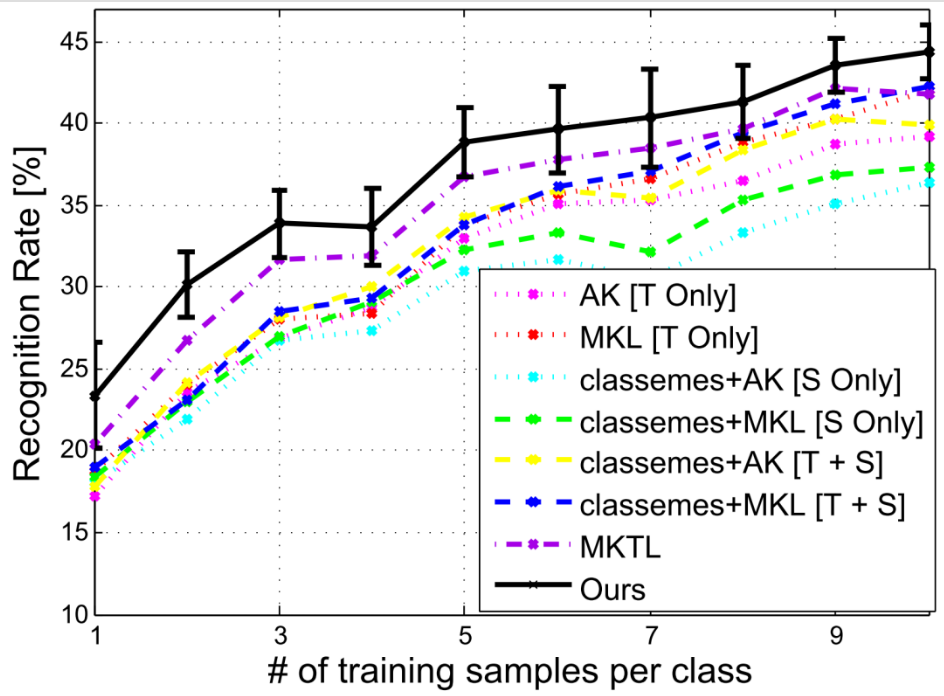
Within-database Transfer Learning

- Recognition rate w.r.t # of training sample per class

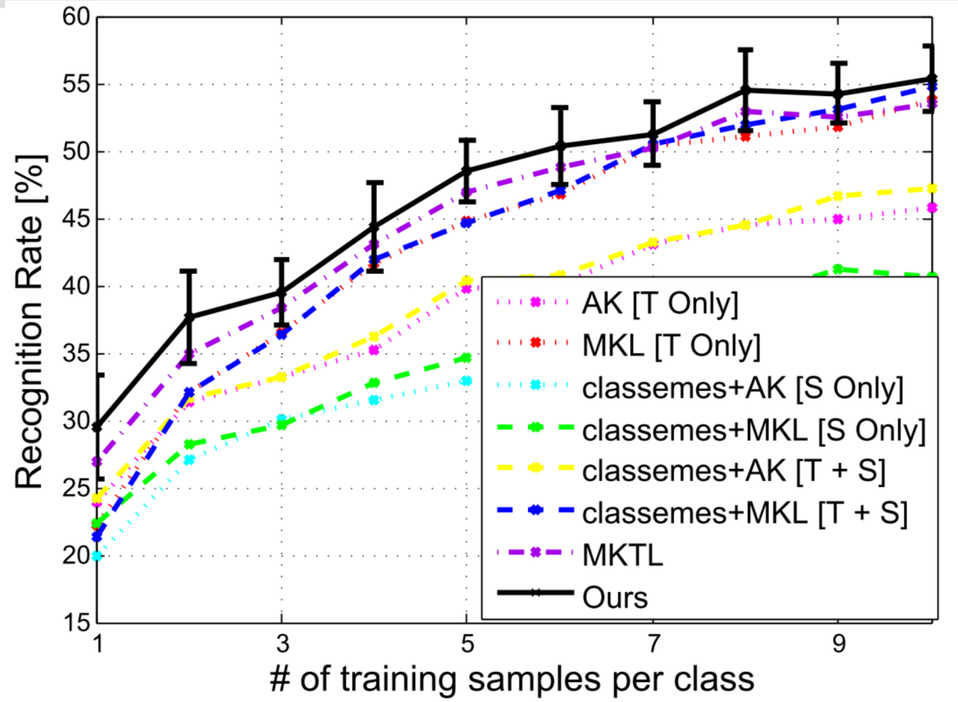


(a) SUN09 (b) Caltech256

Cross-database Transfer Learning



Source: MSRC, Target: SUN09



Source: MSRC, Target: Caltech256

Conclusions

- The proposed method transfers knowledge from multiple classes to multiple classes via two-layer boosting architecture

- No assumptions about the relatedness of the source and target domains are made

- Knowledge transfer via attributes:
  - Discriminant, learnable, and complementary

- The proposed framework can be extended to zero-shot learning
  - Learn a new object category without training data

Thank You