Lecture slides available at

http://goo.gl/MdA6vi

台灣人工智慧學校技術領袖培訓班

Transfer Learning:

Part 2. Challenges in Transfer Learning

Yu-Chiang Frank Wang 王鈺強, Associate Professor

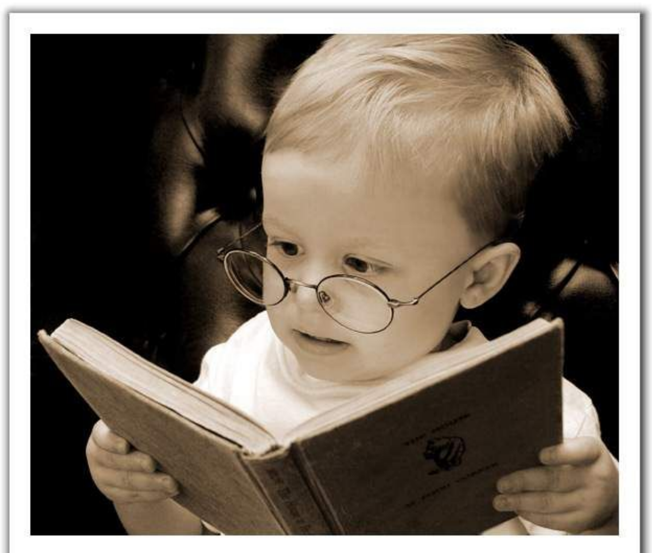
Graduate Inst. Comm. Engineering & Dept. Electrical Engineering

National Taiwan University

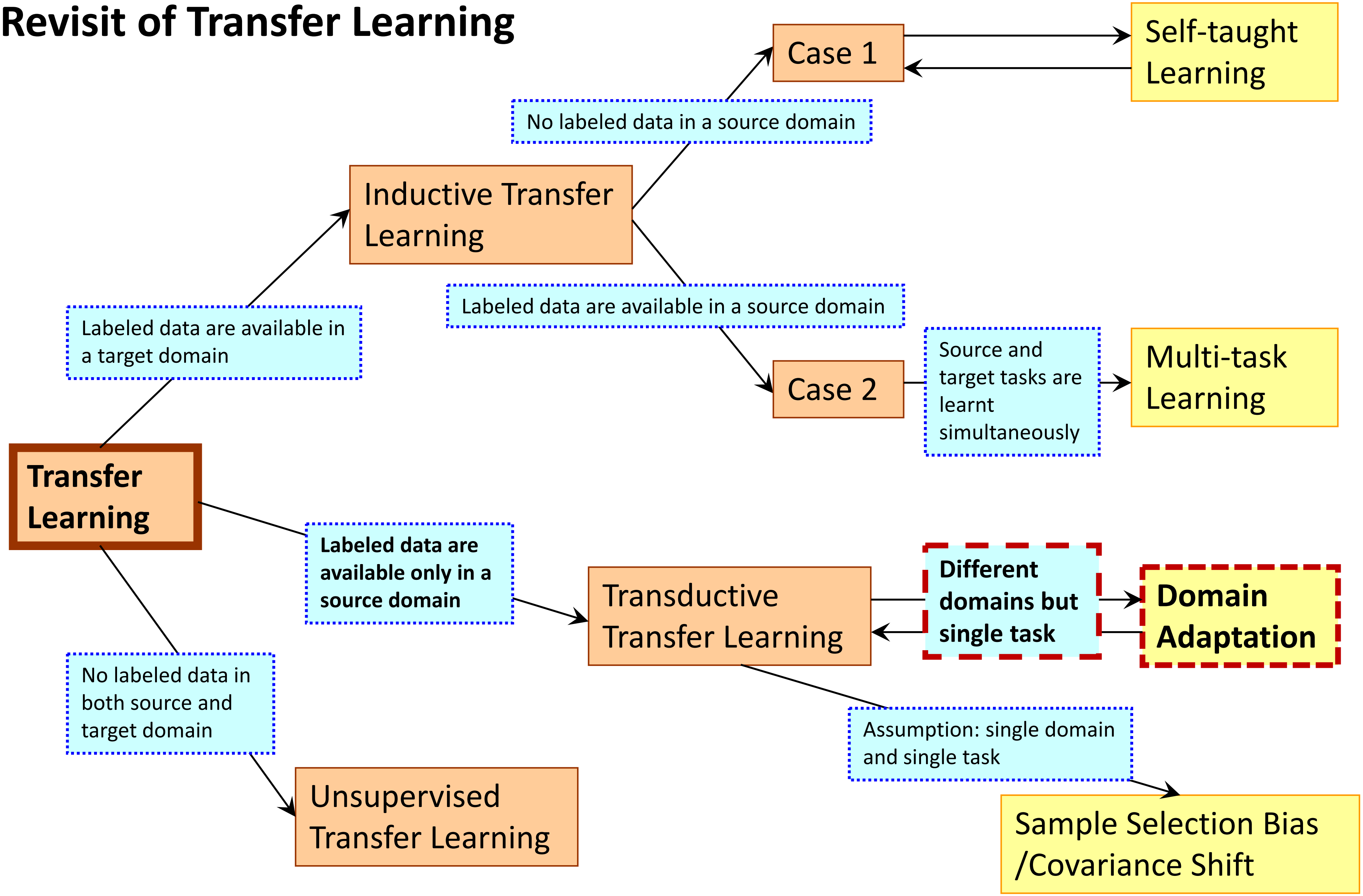
2018/02/26

Topic #2 (10:50~12:30)

- Transfer Learning
  - Introduction to Transfer Learning (TL)
  - Challenges in Transfer Learning
  - TL for Visual Analysis
  - TL for Visual Synthesis and Manipulation



Revisit of Transfer Learning



S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE TKDE, 2010.

Domain Adaptation in Transfer Learning

- Recall that, DA solves the same learning task across data domains, which would probably be of most interest for a variety of applications.
- Thus, in Part II, we will focus on Domain Adaptation in TL.


《细胞》重磅：中国科学家研发的AI影像诊断系统首登顶级期刊，可诊断眼病和肺炎两大类疾病，准确性匹敌顶尖医生 | 科学大发现

原创 2018-02-23 BioTalker 奇点网

今天，由张康教授领导的广州妇女儿童医疗中心和加州大学圣迭戈分校团队，在顶级期刊《细胞》上发表了一篇人工智能（AI）在医疗领域应用的重磅研究成果：基于深度学习开发出一个能诊断眼病和肺炎两大类疾病的AI系统[1]，准确性匹敌顶尖医生。

这不仅是中国研究团队首次在顶级生物医学杂志发表有关医学人工智能的研究成果；也是世界范围内首次使用如此庞大的标注好的高质量数据进行迁移学习，并取得高度精确的诊断结果，达到匹敌甚至超越人类医生的准确性；还是全世界首次实现用AI精确推荐治疗手段。

《细胞》以封面文章的形式，刊登了来自中国的这项AI研究。

Benchmark Datasets

- Amazon Review Dataset (AMT)
  - Product reviews in different domains
  - Kitchen (K), DVD (D), books (B), and electronics (E)
  - 2 classes, about 5,000 documents for each
  - TFIDF (term frequency–inverse document frequency) feature extracted from processed text.

Blitzer et al., Domain adaptation with coupled subspaces, AIS11

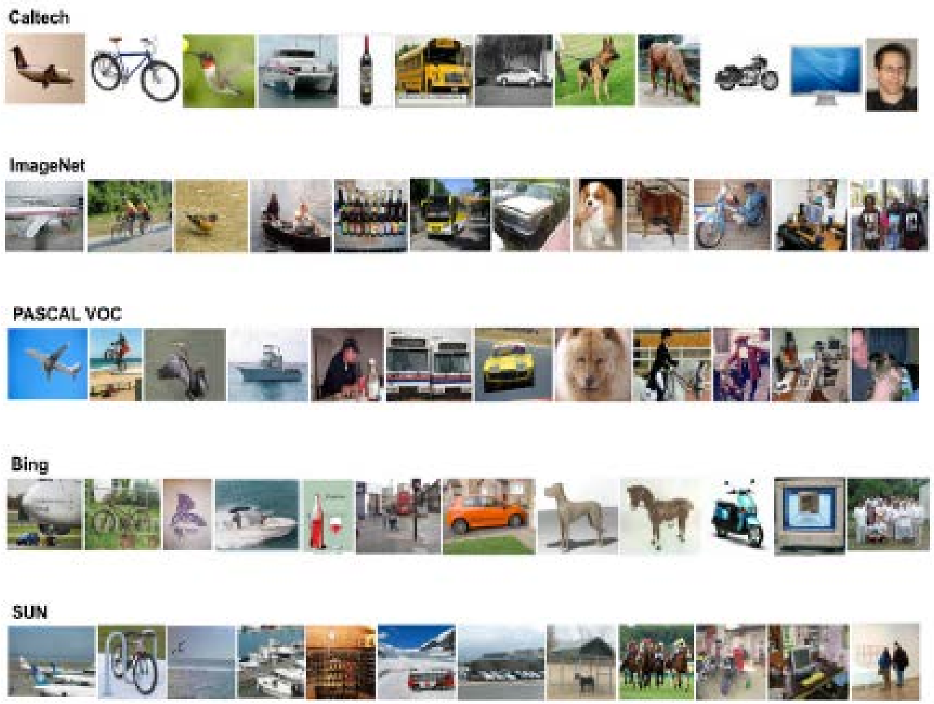
Benchmark Datasets (cont'd)

- Office31 & Office+Caltech
  - Object recognition
  - Amazon (A), Caltech (C), DSLR (D), Webcam (W)
  - 3 domains and 31 classes in Office31
  - 4 domains and 10 common classes in Office+Caltech
  - SURF BoW (Bag-of-Words) and DeCaf6 CNN features are extracted.

Saenko et al., Adapting visual category models to new domains, ECCV10
Gong et al., Reshaping visual datasets for domain adaptation, NIPS13

Benchmark Datasets (cont'd)

- The ImageCLEF'14 DA Challenge (ICDA)
  - Object recognition
  - Caltech (C), ImageNet (I), Pascal (P), Bing (B), SUN (S)
  - 12 classes, about 60 documents for each
  - SIFT BOW features from images



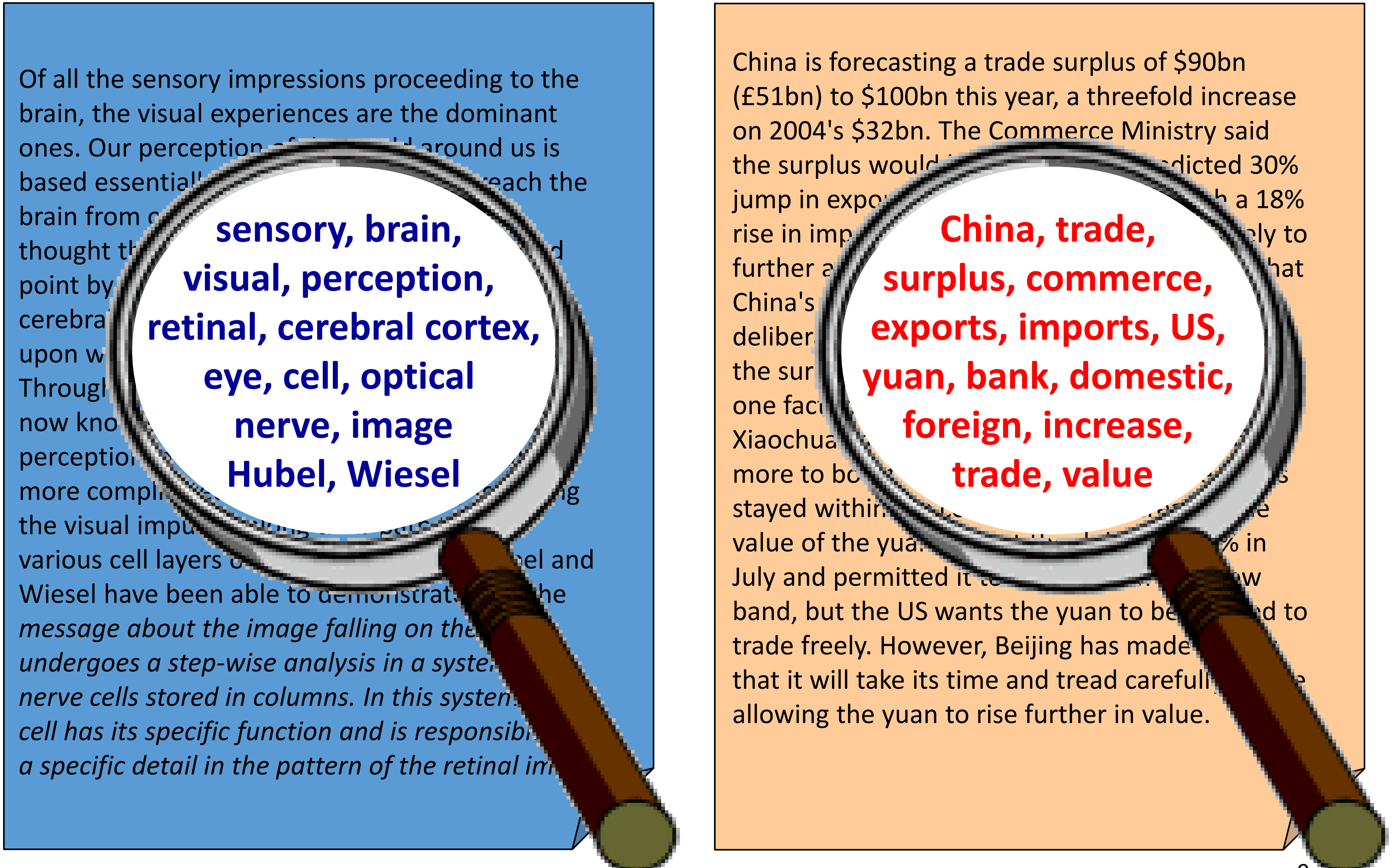
http://www.imageclef.org/2014/adaptation

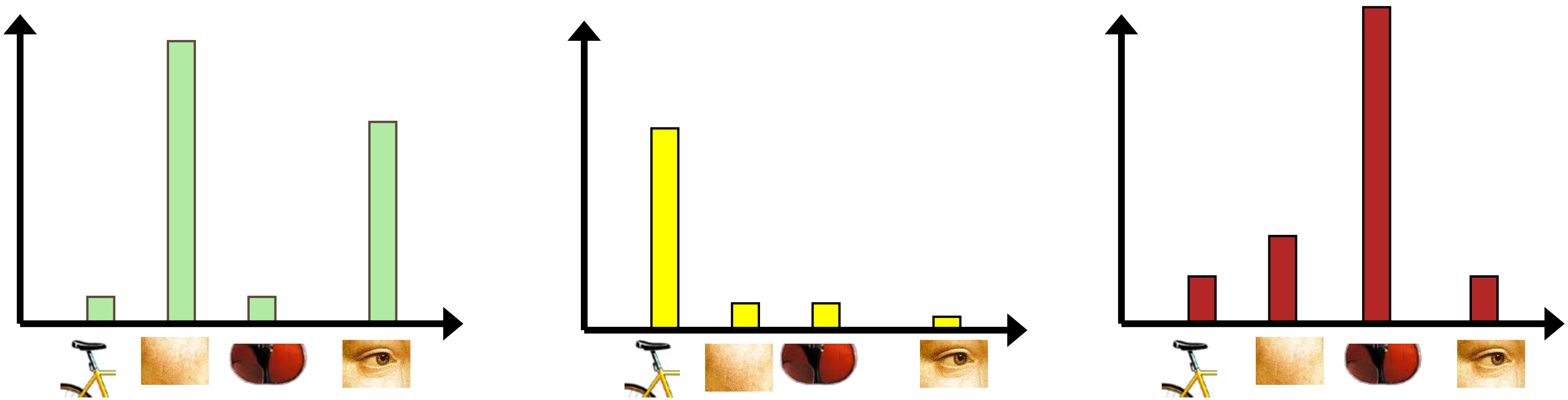
A Very Brief Review of BoW Features

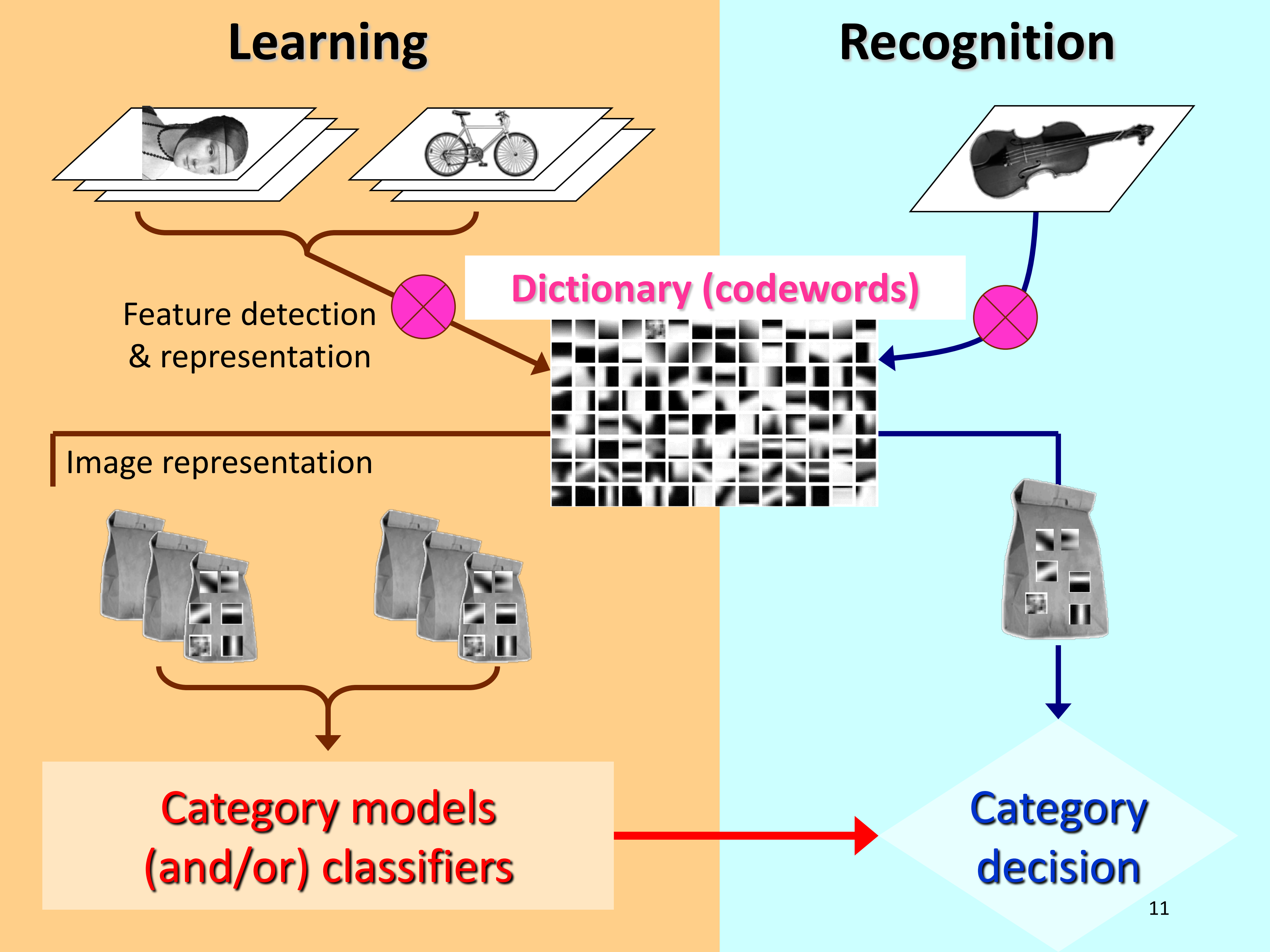
- Before the resurgence of deep learning…

Object Image → **Bag of ‘words’**

BoW: Analogy to Document Representation

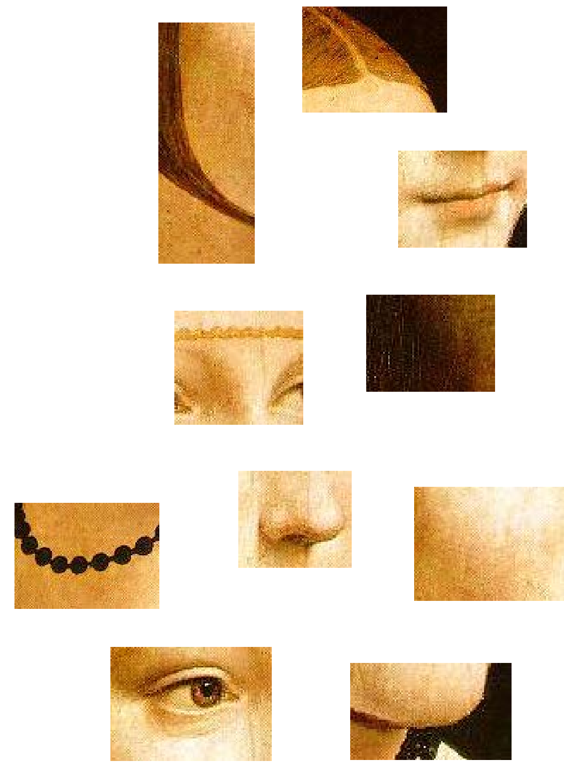




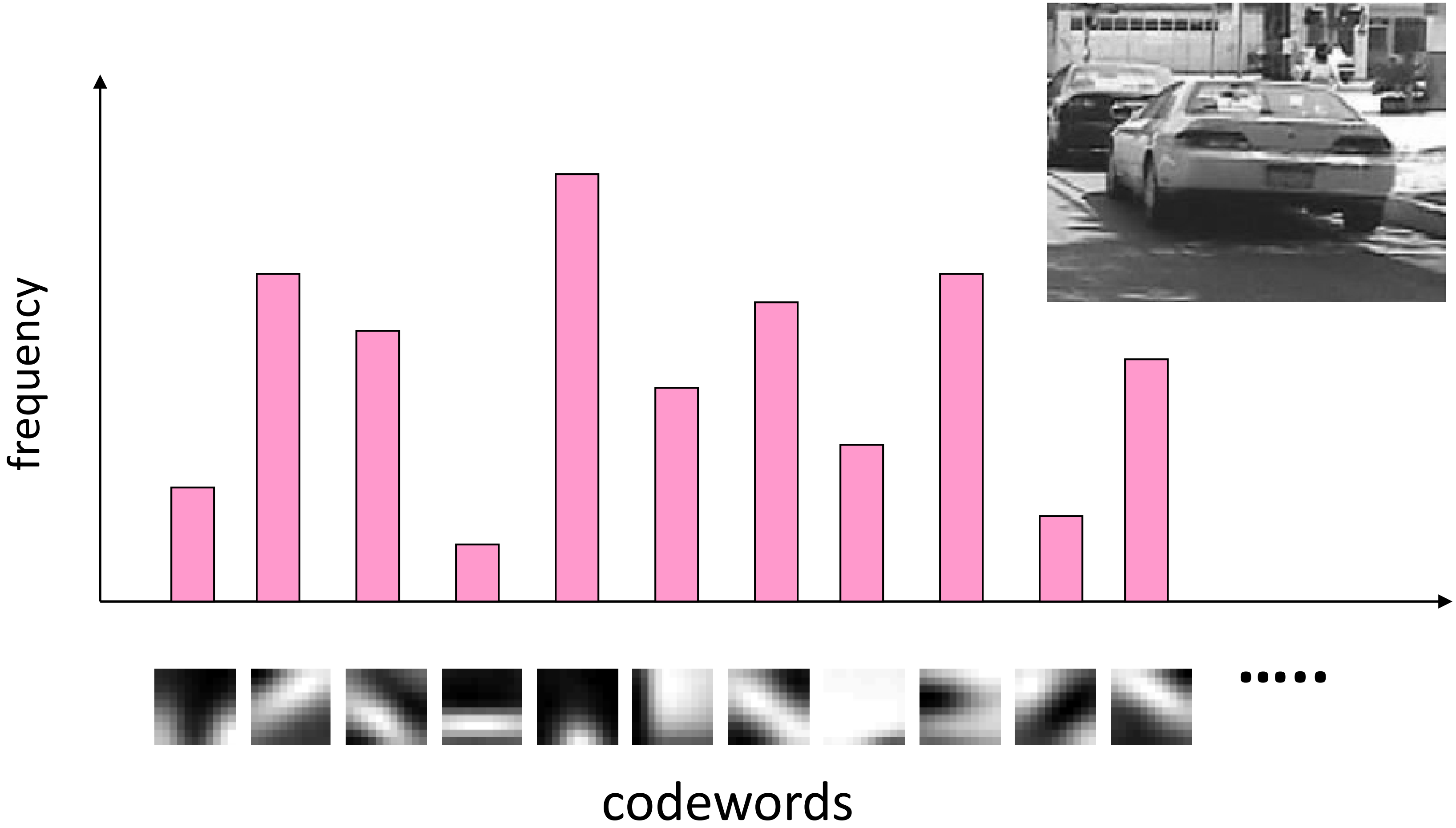


Learning
Recognition
Feature detection
& representation
Dictionary (codewords)
Image representation
Category models
(and/or) classifiers
Category
decision

Feature detection and representation for BoW



- BoW is a histogram of codewords, which counts the occurrences of each codeword in an image.



Challenges in Domain Adaptation

- Domain Shift/Bias/Mismatch
- Data Types: Homogeneous vs. Heterogeneous DA
- Settings for Training DA Models:
  From Supervised, Semi-Supervised to Unsupervised DA

Challenge #1 in Domain Adaptation

- **Domain Shift**
  - AKA *domain bias*, *domain mismatch*, etc.
  - Image classification: different view points, sensors, etc.
  - Audio recognition: different speakers, environments, quality, etc.
  - Activity recognition: different identities, context, etc.
  - Semantic analysis: different topics, vocabularies, etc.

Training data in the source domain

Test data in the target domain

A Popular Technique to Eliminate Domain Shift

- **Maximum Mean Discrepancy**
  - Minimizing the MMD between domains (Huang et al., NIPS'06):

$$MMD(S,T) = \left\| \frac{1}{N_s}\sum_{i=1}^{N_s}\phi(\mathbf{x}_i^s) - \frac{1}{N_t}\sum_{j=1}^{N_t}\phi(\mathbf{x}_j^t)\right\|_{\mathcal{H}}$$

where $\mathcal{H}$ is the RKHS (reproducing kernel Hilbert space) associated with the kernel k, and $\phi(\mathbf{x}) =< k(\mathbf{x}), . >$.

  - Empirically (recall what is done in SVM via kernels):

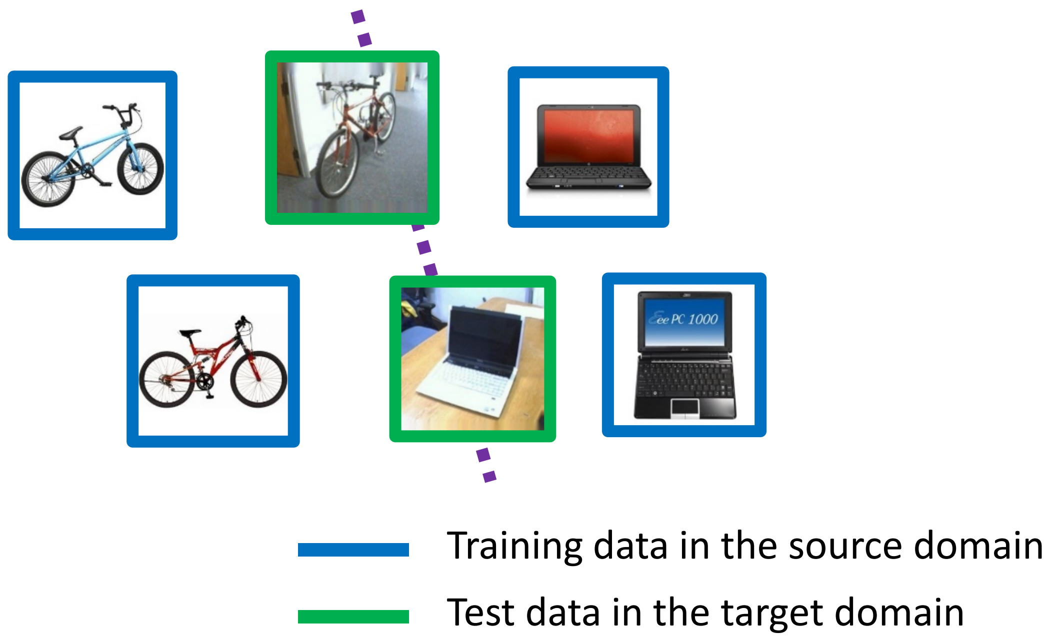
$$MMD(S,T) = \left[\frac{1}{N_s^2}\sum_{i,j=1}^{N_s} k(\mathbf{x}_i^s,\mathbf{x}_j^s) - \frac{2}{N_s N_t}\sum_{i,j=1}^{N_s,N_t} k(\mathbf{x}_i^s,\mathbf{x}_j^t) + \frac{1}{N_t^2}\sum_{i,j=1}^{N_t} k(\mathbf{x}_i^t,\mathbf{x}_j^t)\right]$$

with k being *e.g.* the Gaussian Kernel.

  - Will see some examples in Part III (TL for Visual Analysis).

Challenge #2 in Domain Adaptation

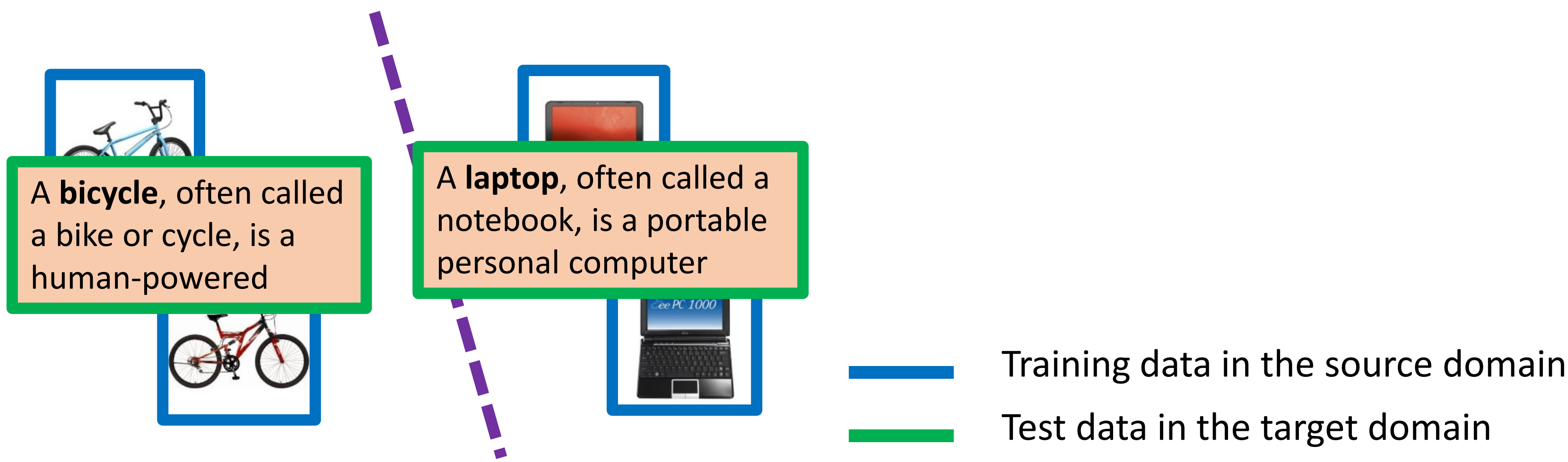
- Data Types: **Homogeneous vs. Heterogeneous DA**
- Homogeneous DA deals with cross-domain data of the same type of feature representations (but with distinct distributions, etc.).



- E.g., action recognition using data captured in different view points, speech recognition using data recorded by different speakers, etc.

Challenge #2 in Domain Adaptation

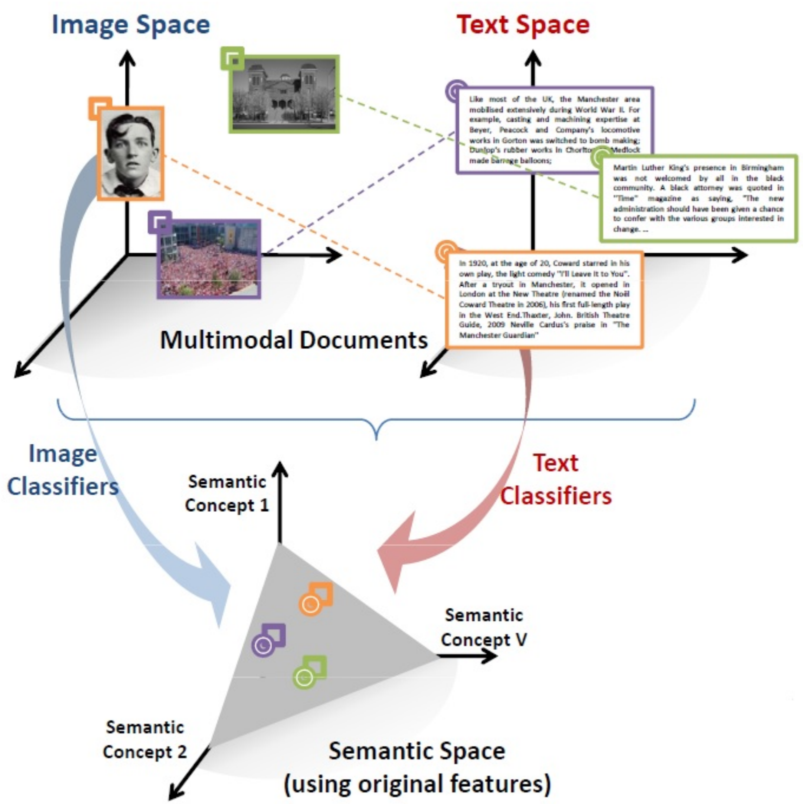
- Data Types: Homogeneous vs. Heterogeneous DA
- Heterogeneous DA deals with cross-domain data of the distinct types of feature representations (and thus with very different distributions).



- When can we expect heterogeneous DA? For example…
  - Image classification with source-domain data in feature 1 but target-domain data in feature 2 (e.g., SIFT/HOG vs. deep features).
  - Image-to-text or text-to-image retrieval/recognition also deal with cross-domain heterogeneous data.

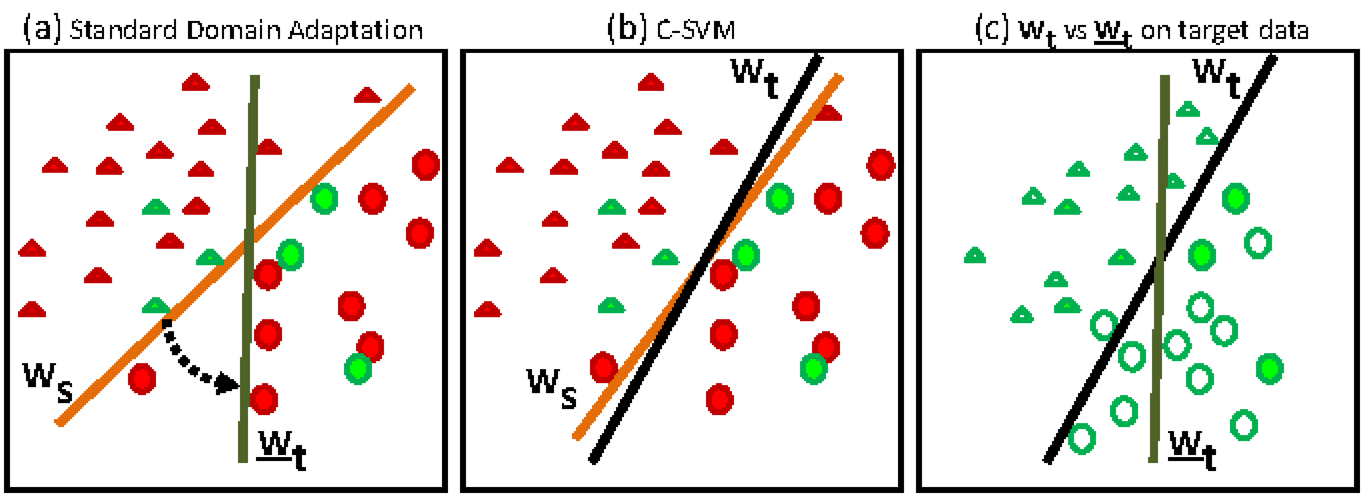
Challenge #2 in Domain Adaptation

- Remarks for Heterogeneous DA
  - Since very different data representations and distributions, one generally, expect at least few labeled data in the target domain.
  - Thus, the setting of semi-supervised DA is preferable (if not required).



Challenge #3 in Domain Adaptation

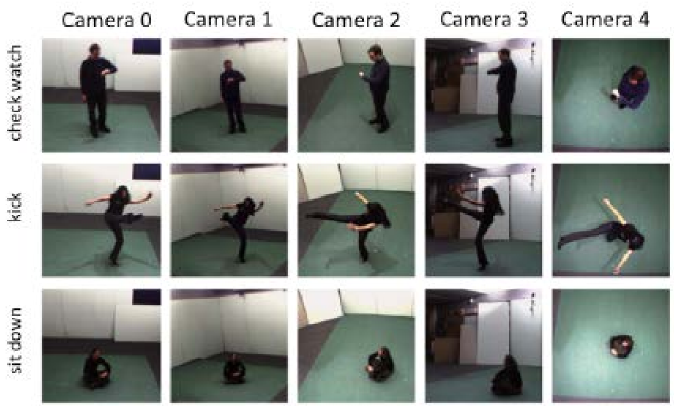
- Settings for Training DA Models:
  From Supervised, Semi-Supervised to Unsupervised DA
- Supervised DA:
  - Both source and target-domain data are with labels during training, which is relatively rare to see (and sometimes might not require DA at all).



Venkateswara et al., Coupled Support Vector Machines for Supervised Domain Adaptation, ACM Multimedia, 2015

Challenge #3 in Domain Adaptation

- Examples of Supervised DA:
  - Cross-domain data pairs are available during training.
  - Commonly seen in applications like person re-identification, cross-camera action recognition or heterogeneous face recognition.



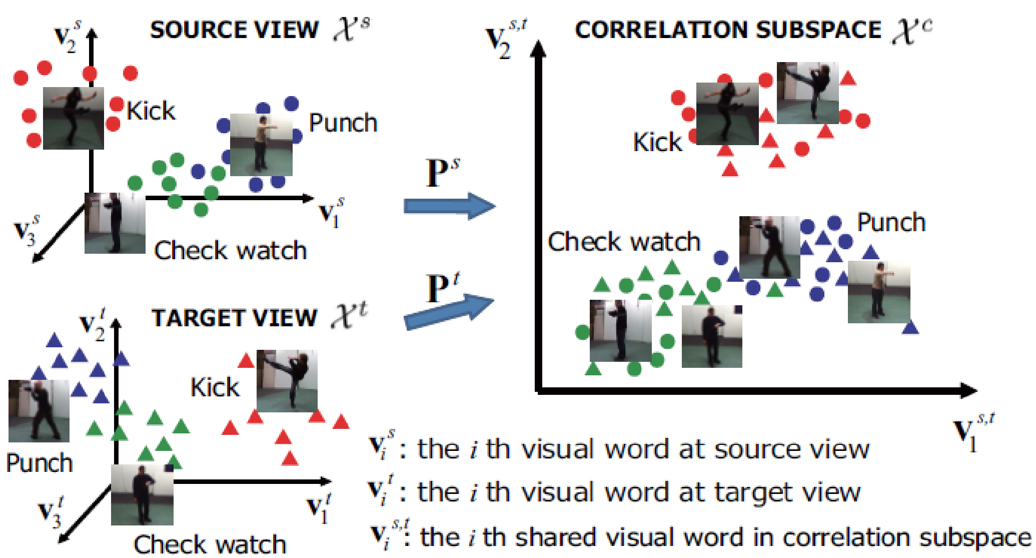
Gray et al., Evaluating appearance models for recognition, reacquisition, and tracking, IEEE PETS Workshop, 2007.
Wang and Tang, Face photo-sketch synthesis and recognition, IEEE PAMI, 2010
Li et al., The CASIA NIR-VIS 2.0 Face Database, PBVS, 2013

Popular Solutions to Supervised DA

- Canonical correlation analysis (CCA)

$$\max_{\mathbf{u}^s,\mathbf{u}^t} \rho = \frac{\mathbf{u}^{s\top}\Sigma_{st}\mathbf{u}^t}{\sqrt{\mathbf{u}^{s\top}\Sigma_{ss}\mathbf{u}^s}\sqrt{\mathbf{u}^{t\top}\Sigma_{tt}\mathbf{u}^t}},$$

where $\Sigma_{st} = \mathbf{X}^s\mathbf{X}^{t\top}$, $\Sigma_{ss} = \mathbf{X}^s\mathbf{X}^{s\top}$, $\Sigma_{tt} = \mathbf{X}^t\mathbf{X}^{t\top}$, and $\rho \in [0, 1]$.



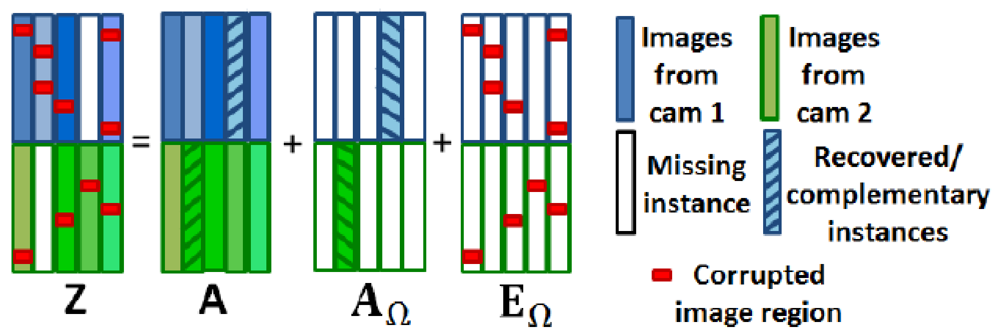
Yeh et al., Heterogeneous Domain Adaptation and Classification by Exploiting the Correlation Subspace, IEEE TKDE, 2014

Popular Solutions to Supervised DA (cont'd)

- Robust PCA or Low-Rank Matrix Decomposition

$$\min_{\mathbf{A},\mathbf{E}_\Omega} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}_\Omega\|_1 \quad s.t. \quad \mathbf{Z} = \mathbf{A} + \mathbf{A}_\Omega + \mathbf{E}_\Omega$$

Z=[**X**; **Y**]: Observed cross-view image data
Ω: Set of observed data
A: Predicted cross-view image data
$\mathbf{A}_\Omega$: Recovered (missing) cross-view data
$\mathbf{E}_\Omega$: Error matrix

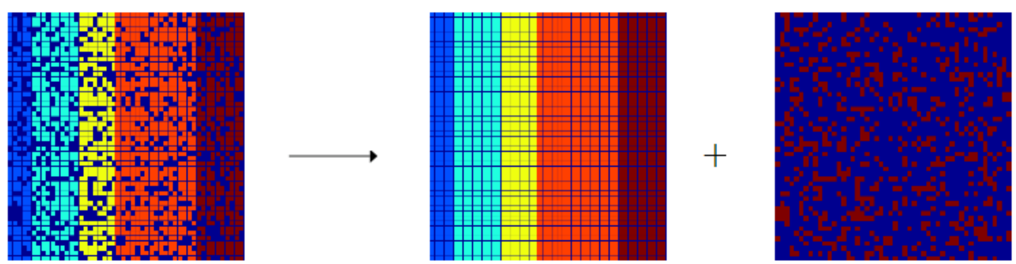


M.-H. Fu, Y.-C. F. Wang, and C.-S. Chen, IEEE ICIP, 2014.

Challenge #3 in Domain Adaptation

- Popular Solutions to Supervised DA:
  - Robust PCA or Low-Rank Matrix Decomposition
- Formulation
  - Given the observed data matrix $\boldsymbol{D}$ as a *m* x *n* matrix, recover a *low-rank* matrix $\boldsymbol{A}$, which satisfies $\boldsymbol{D} = \boldsymbol{A} + \boldsymbol{E}$ while $\boldsymbol{E}$ is a sparse matrix.
  - Also known as Robust PCA
  - Objective function:

$$\min rank\,(\boldsymbol{A}) + \gamma \|\boldsymbol{E}\|_0, subject\ to\ \boldsymbol{D} = \boldsymbol{A} + \boldsymbol{E}$$

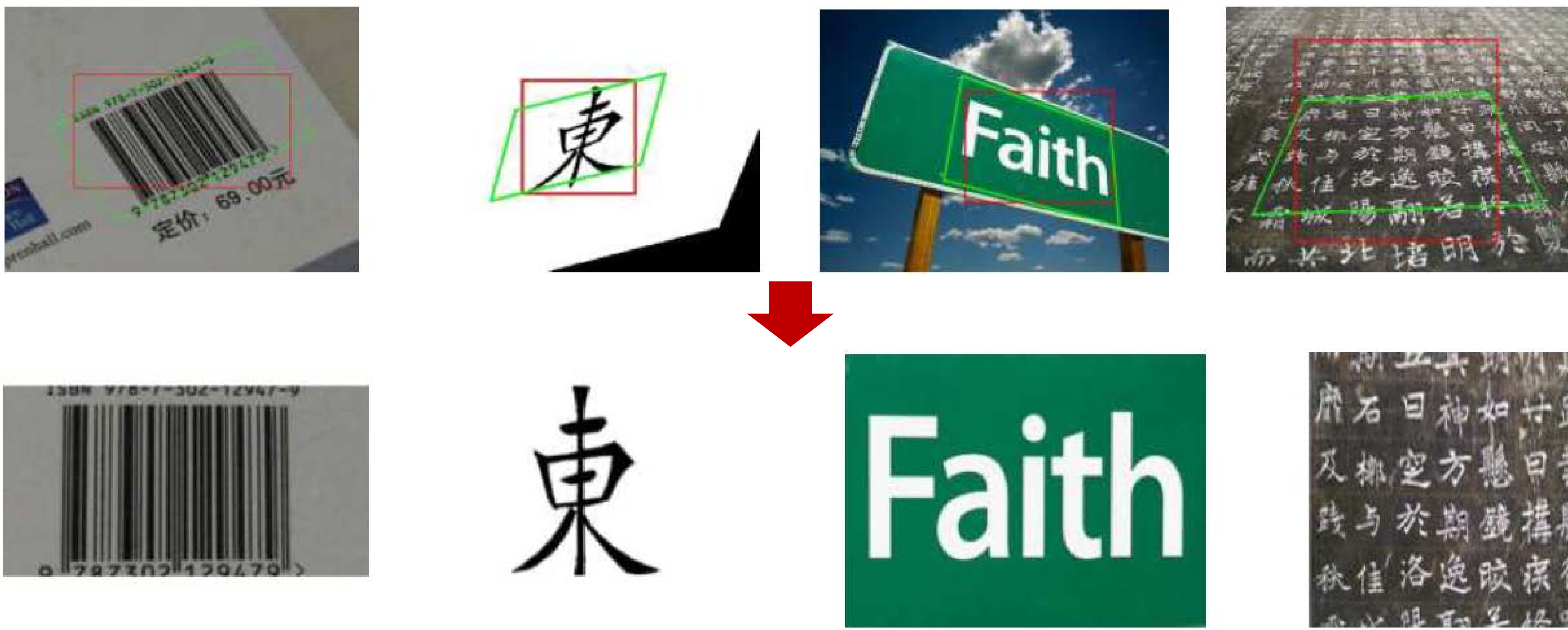


Matrix of corrupted observations $\boldsymbol{D}$ | *Underlying low-rank matrix* $\boldsymbol{A}$ | *Sparse error matrix* $\boldsymbol{E}$

http://matlabstuff.wordpress.com/2009/11/18/78/

Challenge #3 in Domain Adaptation

- Popular Solutions to Supervised DA:
  - More examples of Robust PCA (Low-Rank Matrix Decomposition)



Zhang et al., IJCV'12; slide credit: B. Ghanem

Challenge #3 in Domain Adaptation

- Popular Solutions to Supervised DA:
  - More examples of Robust PCA (Low-Rank Matrix Decomposition)

$$\min_{\mathbf{A},\mathbf{E}} rank(\mathbf{A}) + \lambda \|\mathbf{E}\|_0 \quad \text{s.t.} \quad \mathbf{D} = \mathbf{A} + \mathbf{E}.$$

(a) Original images $\mathbf{D}$

(b) Low-rank and approximated images $\mathbf{A}$ of (a)

(c) Sparse error images $\mathbf{E}$ of (a)

Chen et al., Low-Rank Matrix Recovery with Structural Incoherence for Robust Face Recognition, CVPR 2012

Challenge #3 in Domain Adaptation

- Popular Solutions to Supervised DA:
  - More examples of Robust PCA (Low-Rank Matrix Decomposition)

$$\min_{\mathbf{A},\mathbf{E}} rank(\mathbf{A}) + \lambda \left\|\mathbf{E}\right\|_0 \quad \text{s.t.} \;\; \mathbf{D} = \mathbf{A} + \mathbf{E}.$$

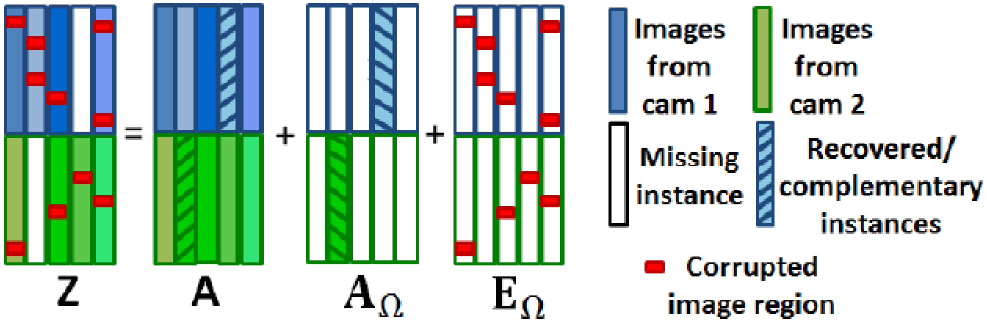
E. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? Journal of the ACM, 58, 2011

Challenge #3 in Domain Adaptation

- Popular Solutions to Supervised DA:
  - Robust PCA/Low-Rank Matrix Decomposition

$$\min_{\mathbf{A},\mathbf{E}_\Omega} \|\mathbf{A}\|_* + \lambda\|\mathbf{E}_\Omega\|_1 \quad s.t. \quad \mathbf{Z} = \mathbf{A} + \mathbf{A}_\Omega + \mathbf{E}_\Omega$$

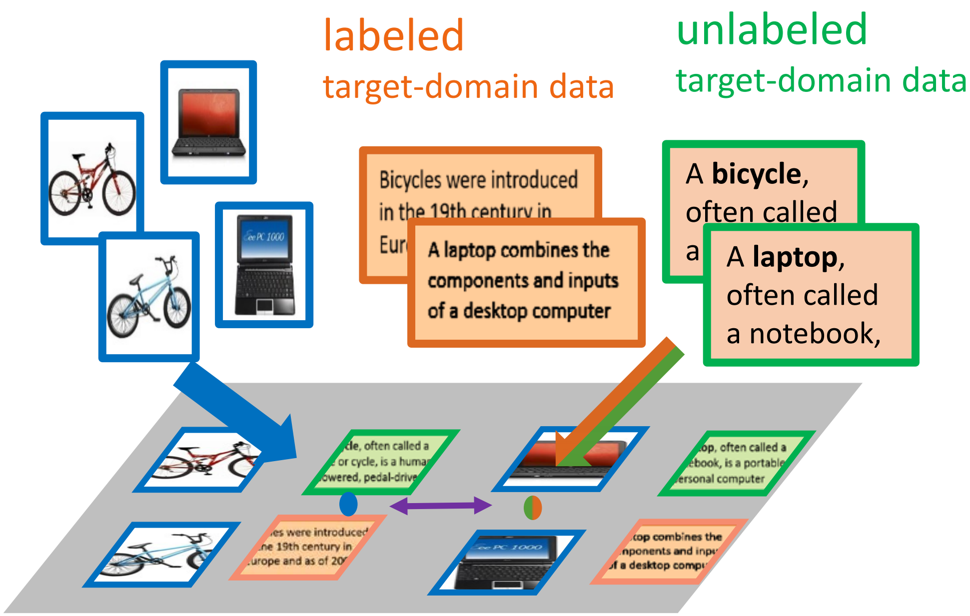
Z=[**X**; **Y**]: Observed cross-view image data
Ω: Set of observed data
A: Predicted cross-view image data
$\mathbf{A}_\Omega$: Recovered (missing) cross-view data
$\mathbf{E}_\Omega$: Error matrix



M.-H. Fu, Y.-C. F. Wang, and C.-S. Chen, IEEE ICIP, 2014.

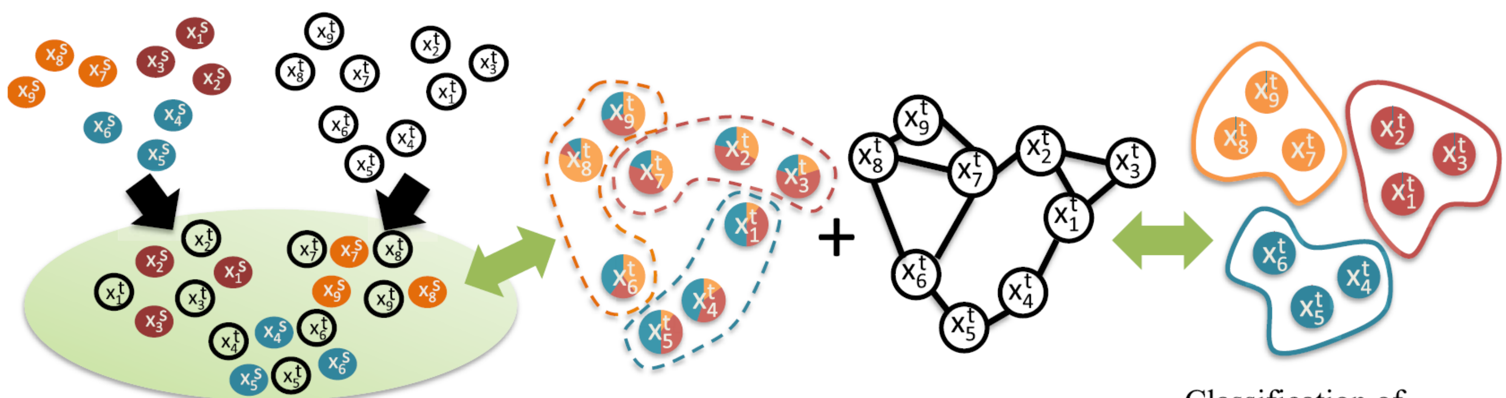
Challenge #3 in Domain Adaptation

- Settings for Training DA Models:
  From Supervised, Semi-Supervised to Unsupervised DA
- Semi-Supervised DA:
  - Source-domain data are fully labeled.
  - Only a small amount of target-domain data are with label info.
  - More practical, and typically seen in real-world applications.
- Examples



Challenge #3 in Domain Adaptation

- Settings for Training DA Models:
  From Supervised, Semi-Supervised to Unsupervised DA
- Unsupervised DA:
  - Source-domain data are fully labeled.
  - No label information is available in target domain.
  - Also practical, and might benefit a large number of real-world applications.
- Examples



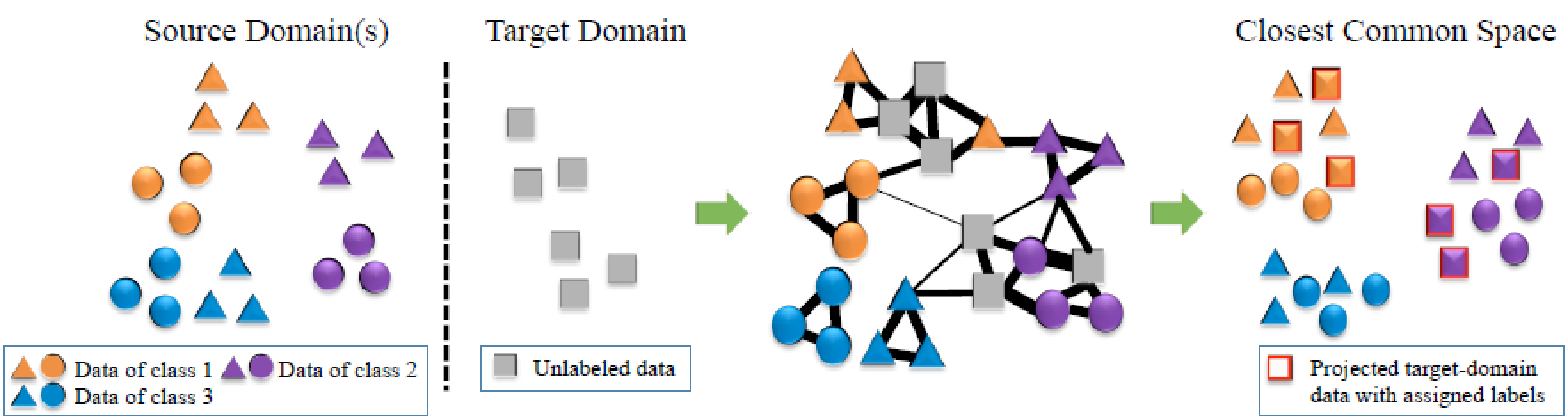
Distribution Adaptation

Exploiting Label & Structural Consistency

Classification of Target-Domain Data

Challenge #3.5 in Domain Adaptation

- Imbalanced & Unsupervised Domain Adaptation
  - Source-domain data are fully labeled.
  - Possibly more than one source domain is available.
  - No label information is available in target domain.
  - Imbalanced class numbers across domains.
  - Practical and very challenging!



Hsu et al., Unsupervised domain adaptation with imbalanced cross-domain data, IEEE ICCV 2015.

Highlight on Recent Approaches for DA

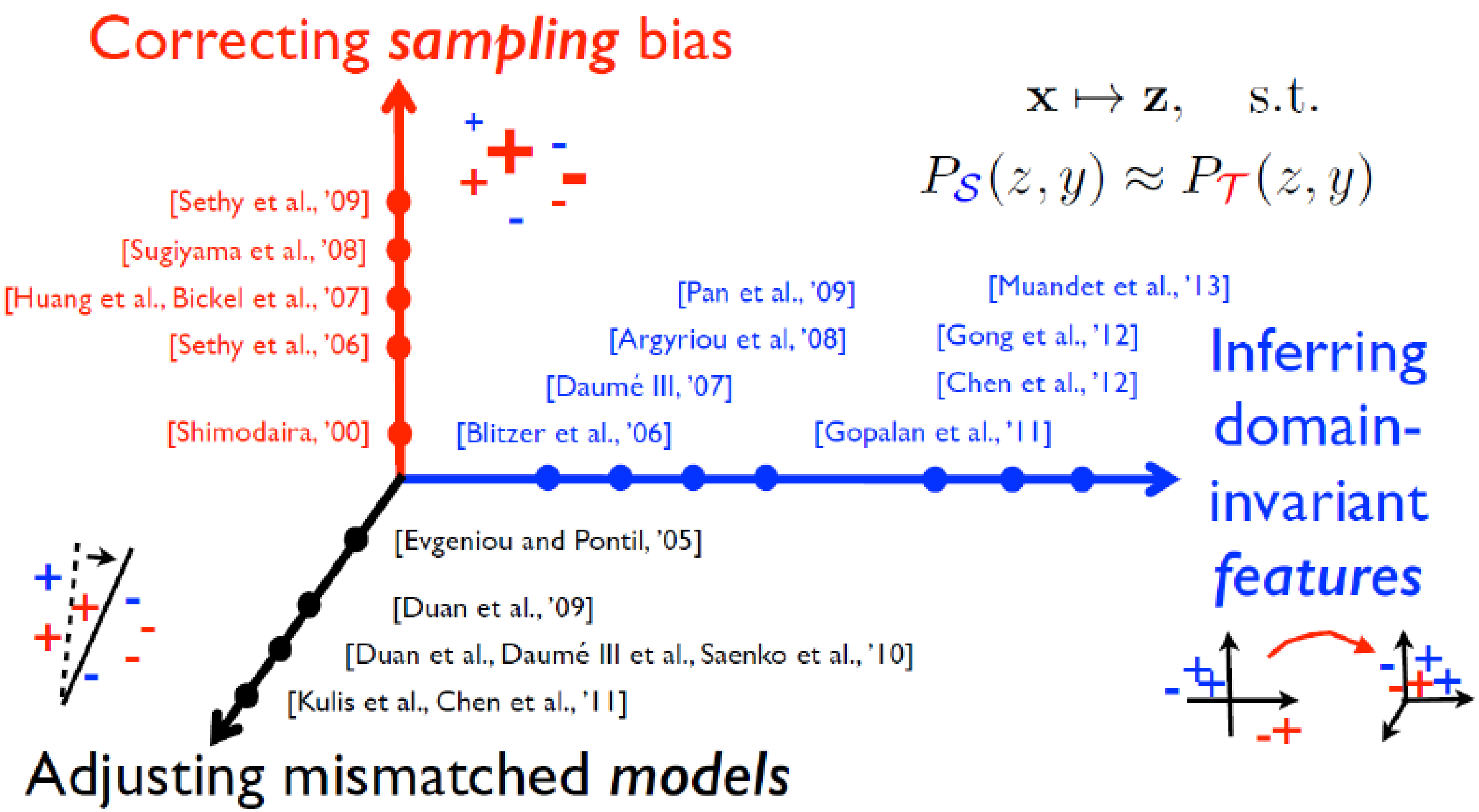
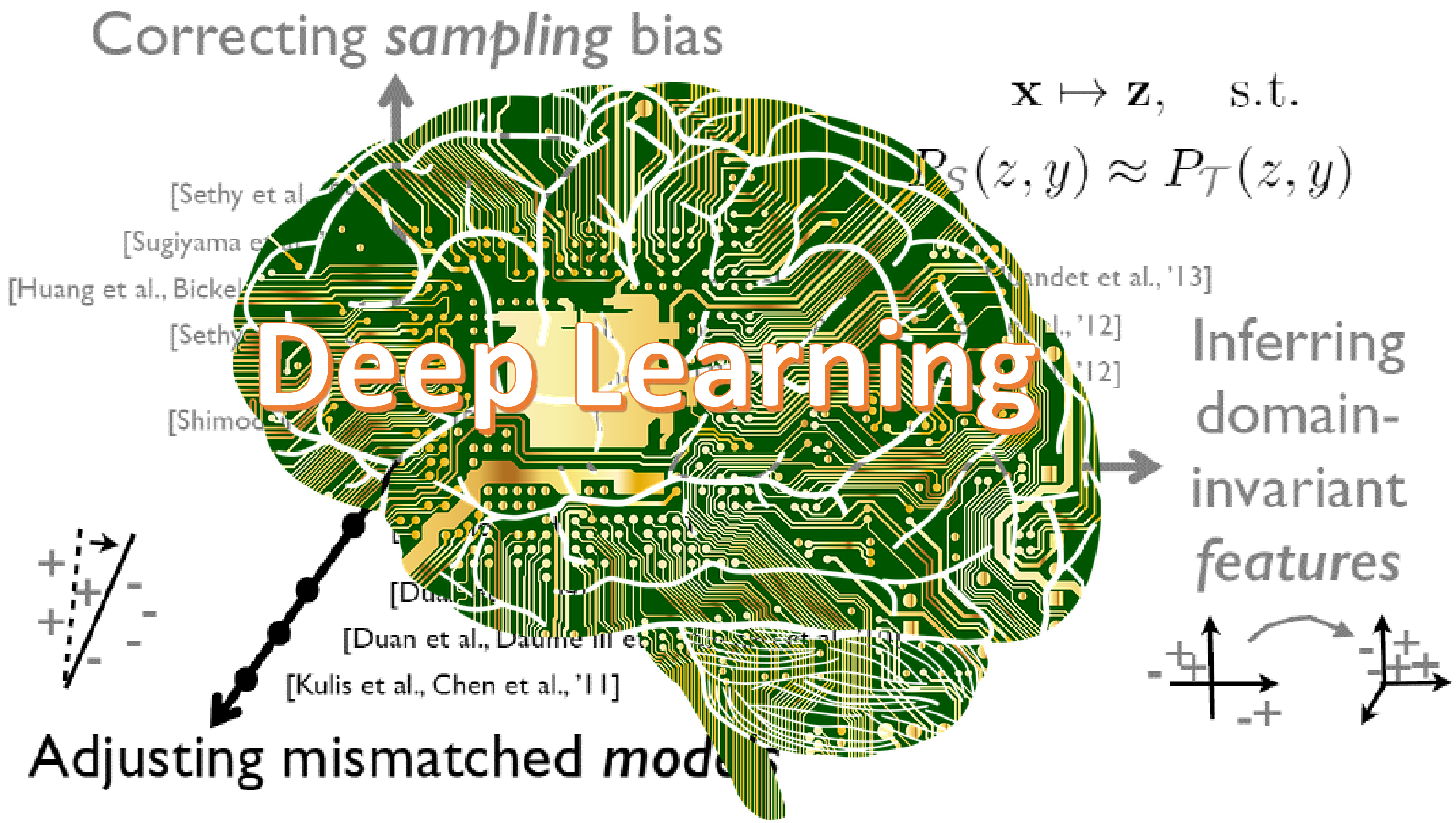


Image: Courtesy to Boqing Gong.

Highlight on Recent Approaches for DA



Let’s Take a Lunch Break…

- Transfer Learning
  - Introduction to Transfer Learning (TL)
  - Challenges in Transfer Learning
  - TL for Visual Analysis
  - TL for Visual Synthesis and Manipulation