台灣人工智慧學校技術領袖培訓班

Transfer Learning:
Part 3. Transfer Learning for Visual Analysis

Yu-Chiang Frank Wang 王鈺強, Associate Professor

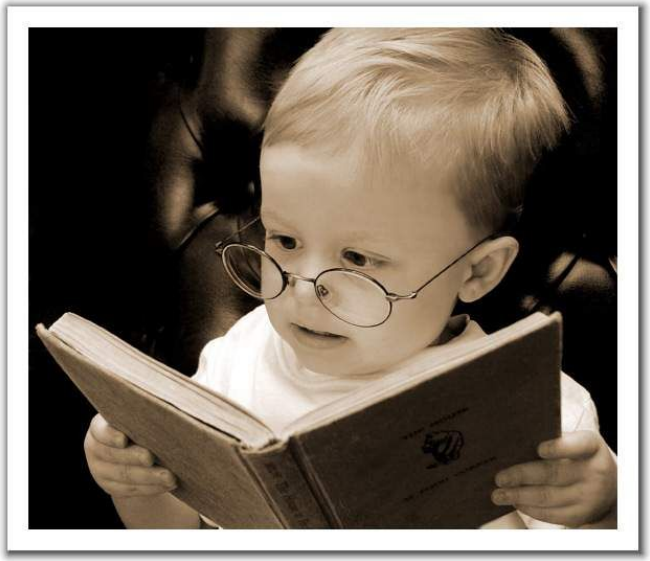
Graduate Inst. Comm. Engineering & Dept. Electrical Engineering

National Taiwan University

2018/02/26

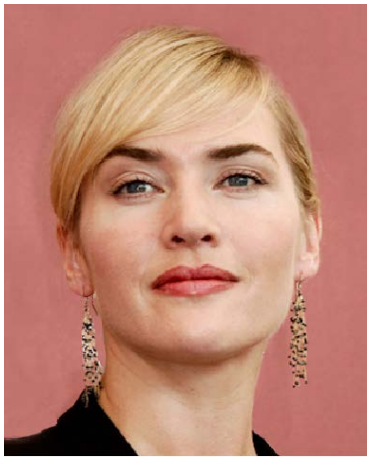
Topic #3 (14:00~15:00)

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



Transfer Learning for Visual Classification

- Recall that
  - Object Recognition

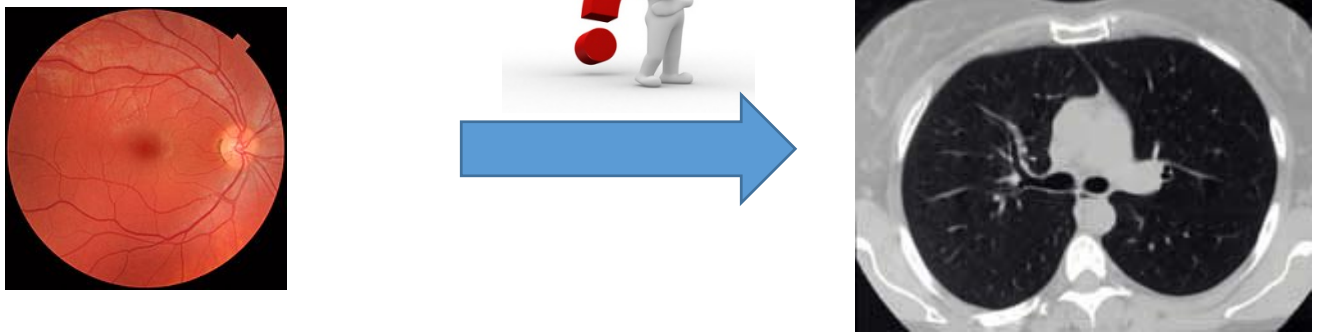


Source Domain

Target Domain

Transfer Learning for Visual Classification (cont'd)

- Recall that
  - Medical Image Classification



Source Domain

Target Domain

Transfer Learning for Visual Classification (cont'd)

- Recall that
  - Application-Oriented Image Classification

Source Domain

Target Domain

TL Approaches for Cross-Domain Classification

- Instance Transfer
  - Re-weight source-domain label instances for adaptation
- Feature Transfer
  - Derive common feature representation for describing cross-domain data
- Parameter Transfer
  - Discover shared learning model parameters for cross-domain data
- Relational Knowledge Transfer (Few/One/Zero-Shot Learning…)
  - Build mapping of relational knowledge between cross-domain data

Methods	Inductive Transfer Learning	Transductive Transfer Learning	Unsupervised Transfer Learning
Instance Transfer (Instance Reweighting)	O	O	
Feature Transfer (Common Feature Representation)	O	O	O
Parameter/Model Transfer	O		
Relational knowledge Transfer	O		

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

Instance Transfer/Reweighting

- Remarks
  - Re-weight source-domain label instances for adaptation
  - Can be viewed as instance selection (not feature selection) in machine learning

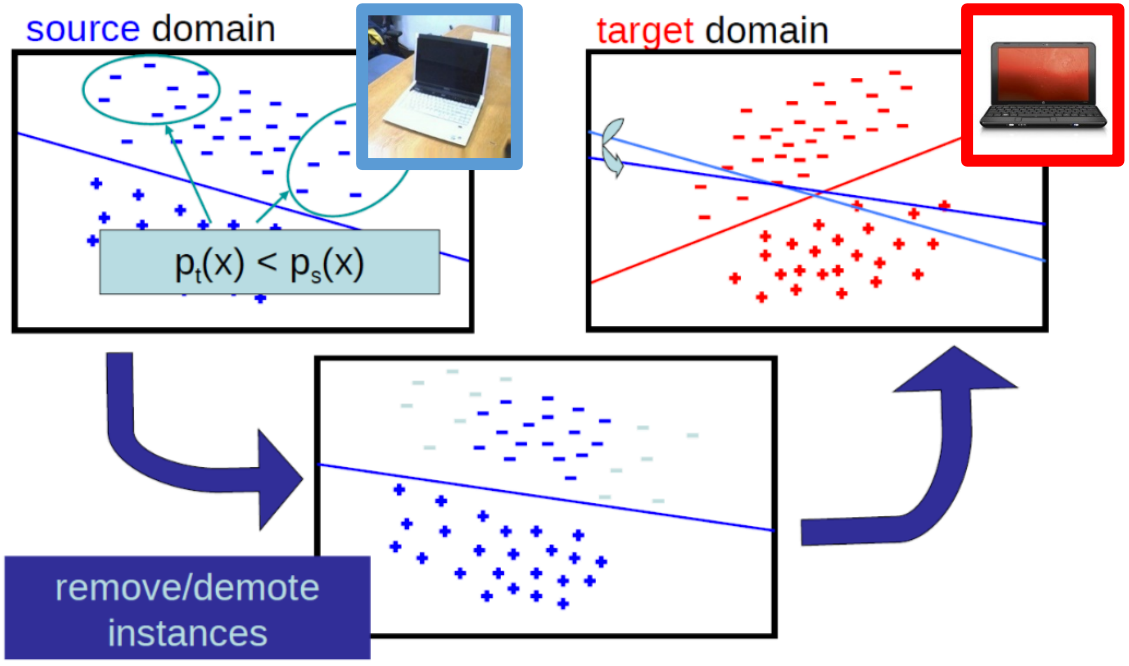


Image: Courtesy to Ming-Wei Chang.

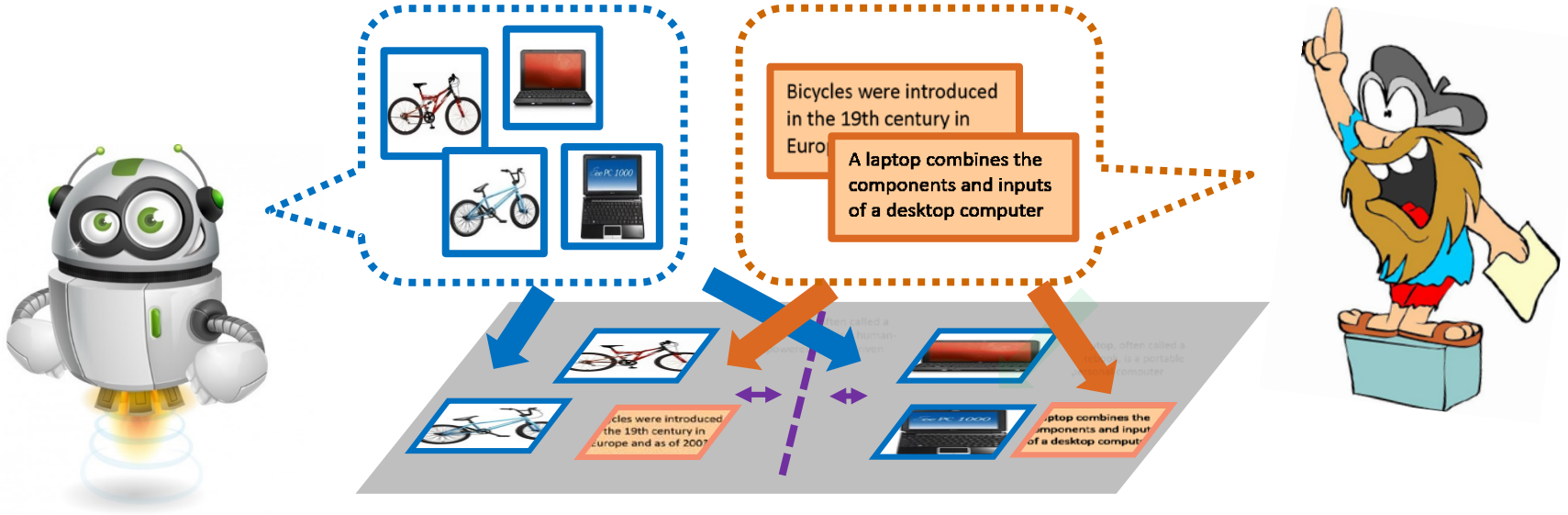
Sugiyama *et al*. NIPS'07

Bickel *et al*. ICML'07

Kanamori *et al*. JMLR'09

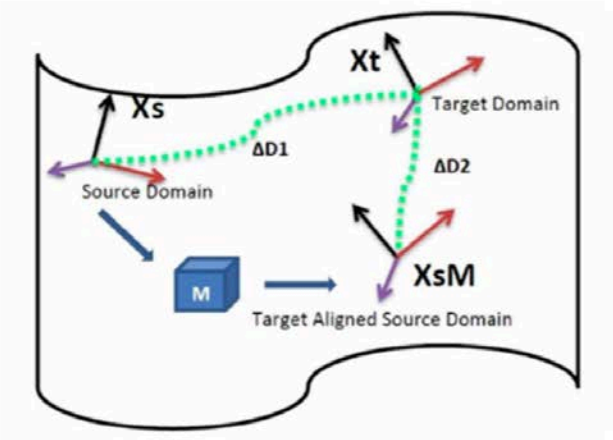
Feature Transfer

- Remarks
  - Learning common feature space for describing cross-domain data
  - Would be preferable if label info can be partially observed in the target domain (i.e., from unsupervised to semi-supervised domain adaptation)
  - When the above remark might become necessary?
    (Hint: homogeneous vs. heterogeneous domain adaptation)



Feature Transfer (cont'd)

- Subspace Alignment
  - Project source-domain data to the target domain (or the reduced dimension version), so that the subspace is properly aligned.
  - Applicable for unsupervised domain adaptation



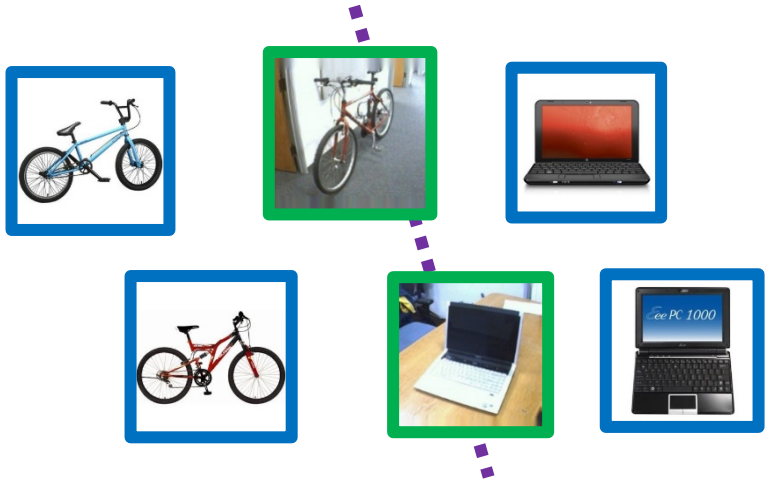
- $\mathbf{M}^* = \mathbf{S}_1'\mathbf{S}_2$ corresponds to the "subspace alignment matrix": $\mathbf{M}^* = \text{argmin}_{\mathbf{M}} \|\mathbf{S}_1\mathbf{M} - \mathbf{S}_2\|$
- $X_a = \mathbf{S}_1\mathbf{S}_1'\mathbf{S}_2 = \mathbf{S}_1\mathbf{M}^*$ projects the source data to the target subspace
- A natural similarity: $\text{Sim}(\mathbf{x}_s, \mathbf{x}_t) = \mathbf{x}_s\mathbf{S}_1\mathbf{M}^*\mathbf{S}_1'\mathbf{x}_t' = \mathbf{x}_s\mathbf{A}\mathbf{x}_t'$

Image: Courtesy to Fernando.

Fernando et al. Unsupervised Visual Domain Adaptation Using Subspace Alignment, ICCV, 2013

Revisit of 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 $\phi(\mathbf{x}) =< k(\mathbf{x}), . >$.

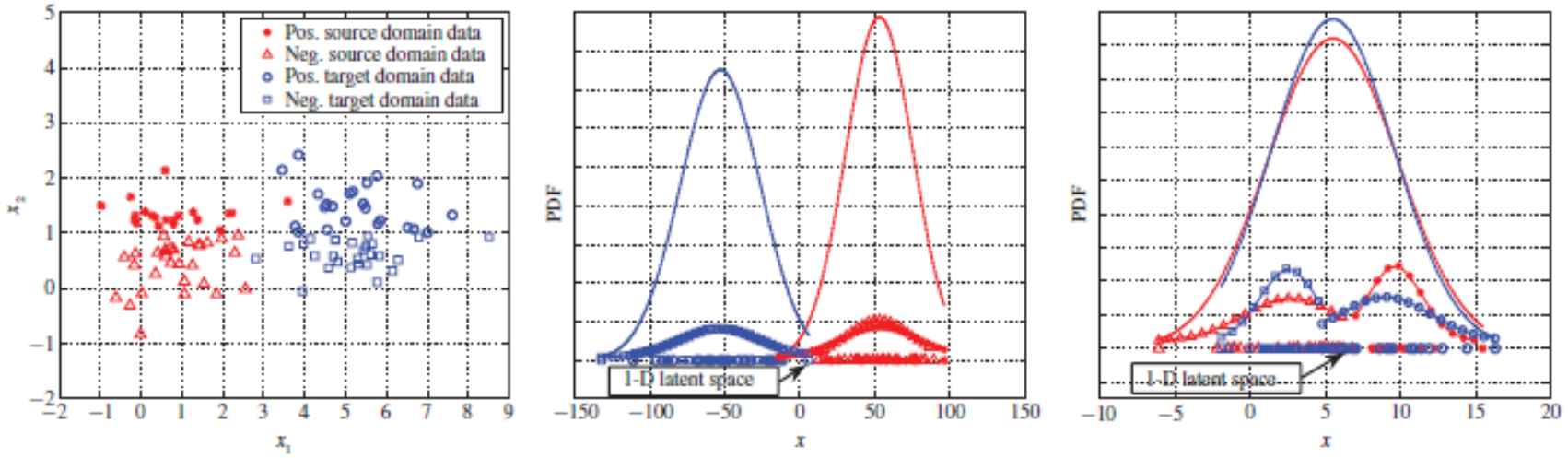
  - Empirically (recall what is done in SVM via kernels):

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

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

Feature Transfer (cont'd)

- Transfer Component Analysis (TCA)
  - Unsupervised DA
  - Matching cross-domain marginal (global) distributions
  - More specifically, minimize distance between $P(\Phi(X_S))$ and $P(\Phi(X_T))$, while $P(\Phi(X))$ is approximated by the global mean of data in each domain.
  - And, preserve the properties of X_S and X_T by maximizing projected data variances.

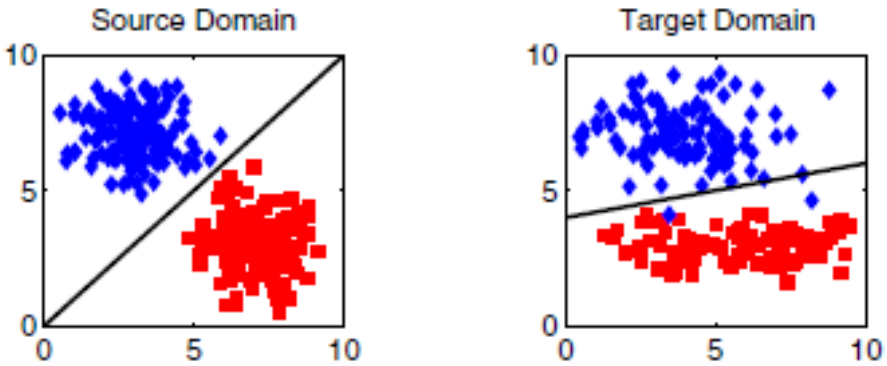


Pan et al., Domain adaptation via transfer component analysis, IEEE Trans. Neural Networks, 2011

Feature Transfer (cont'd)

- Joint Distribution Adaptation (JDA)
  - Unsupervised DA
  - In addition to matching cross-domain marginal (global) distributions, conditional (class-wise) distributions across data domains are also aligned.
  - JDA formulation:

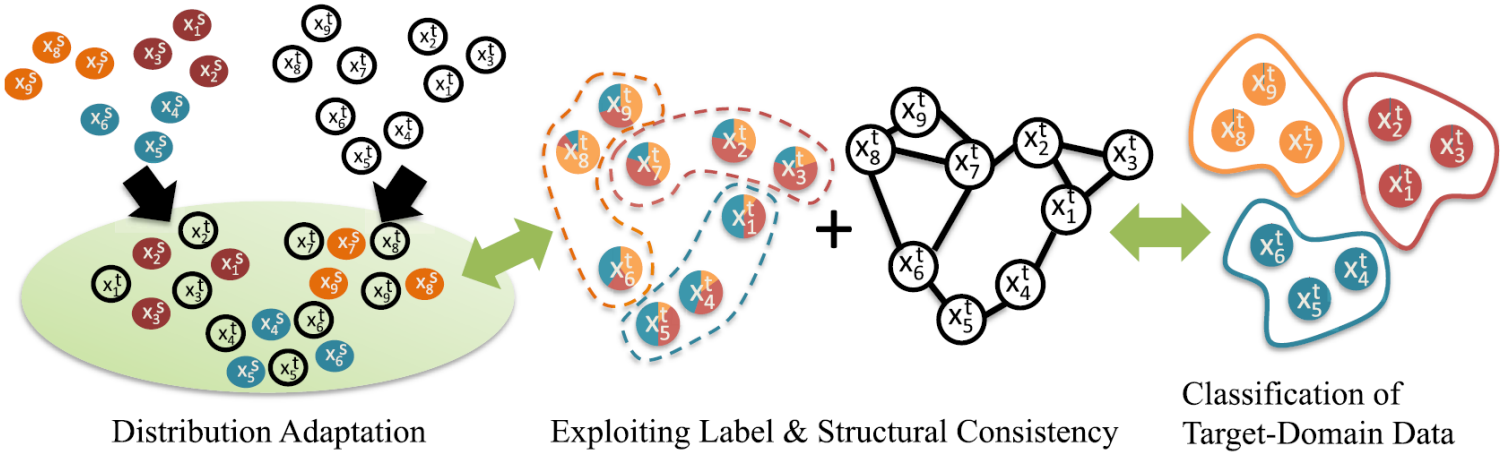
$$\begin{aligned}\min_{T} & \left\|\mathbb{E}_{P(\mathbf{x}_s,y_s)}\left[T\left(\mathbf{x}_s\right),y_s\right]-\mathbb{E}_{P(\mathbf{x}_t,y_t)}\left[T\left(\mathbf{x}_t\right),y_t\right]\right\|^2 \\ \approx & \left\|\mathbb{E}_{P_s(\mathbf{x}_s)}\left[T\left(\mathbf{x}_s\right)\right]-\mathbb{E}_{P_t(\mathbf{x}_t)}\left[T\left(\mathbf{x}_t\right)\right]\right\|^2 \\ + & \left\|\mathbb{E}_{Q_s(y_s|\mathbf{x}_s)}\left[y_s|T\left(\mathbf{x}_s\right)\right]-\mathbb{E}_{Q_t(y_t|\mathbf{x}_t)}\left[y_t|T\left(\mathbf{x}_t\right)\right]\right\|^2\end{aligned}$$



Long et al., Transfer feature learning with joint distribution adaptation, ICCV, 2013

Feature Transfer (cont'd)

- UDA with label and structural consistency
  - Unsupervised DA
  - Similar to JDA, matching of both cross-domain marginal (global) and conditional (class-wise) distributions is desirable.
  - Assign pseudo labels to the target-domain data with confidence + label propagation
  - Bridge the gap between unsupervised and semi-supervised DA.



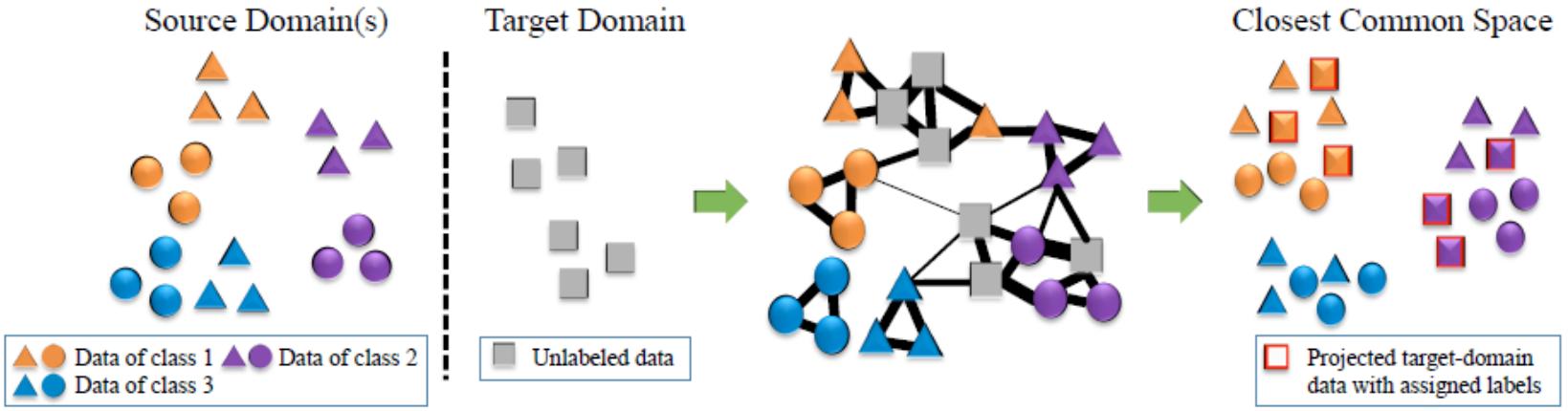
Hou et al., Unsupervised Domain Adaptation With Label and Structural Consistency, IEEE Trans. Image Processing, 2016

Feature Transfer (cont'd)

- Unsupervised + Imbalanced DA
  - Beyond matching cross-domain marginal (global) and conditional (class-wise) distributions by considering label and latent space structural information.

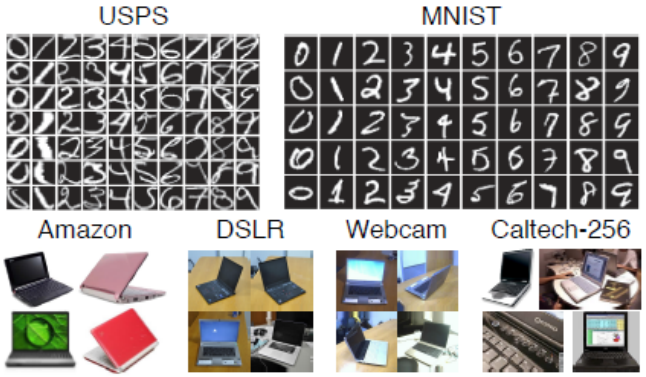
$$\mathcal{M}_\phi\left(\mathcal{P}_S(\mathbf{X}_S,\mathbf{y}_S),\mathcal{P}_T(\mathbf{X}_T,\mathbf{y}_T)\right)$$
$$\approx \mathcal{M}_\phi\left(\mathcal{P}_S(\mathbf{X}_S),\mathcal{P}_T(\mathbf{X}_T)\right)+\mathcal{M}_\phi\left(\mathcal{P}_S(\mathbf{X}_S|\mathbf{y}_S),\mathcal{P}_T(\mathbf{X}_T|\mathbf{y}_T)\right).$$

$$\mathcal{M}_{\phi,d}\left(\mathcal{P}_S(\mathbf{X}_S,\mathbf{y}_S),\mathcal{P}_T(X_T,\mathbf{y}_T)\right)$$
$$\approx \mathcal{M}_\phi\left(\mathcal{P}_S(\mathbf{X}_S),\mathcal{P}_T(\mathbf{X}_T)\right)$$
$$+\mathcal{M}_{\phi,d}\left(\mathcal{P}_S(\mathbf{X}_S|\mathbf{y}_S),\mathcal{P}_T(\mathbf{X}_T|\mathbf{y}_T)\right).$$

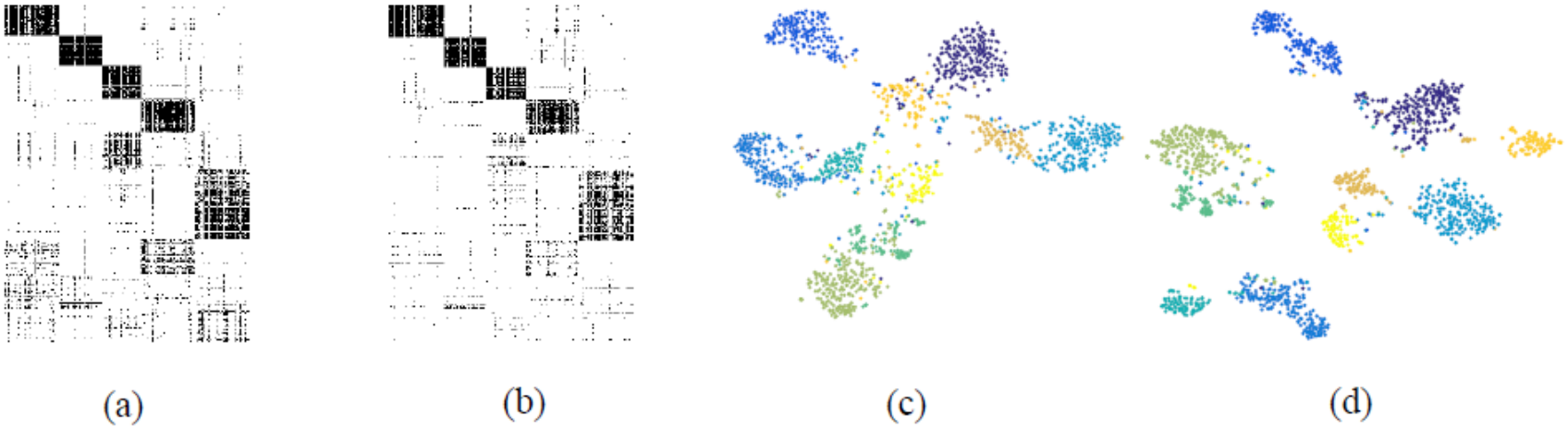


Hsu et al., Unsupervised Domain Adaptation with Imbalanced Cross-Domain Data, ICCV, 2015

Feature Transfer (cont'd)



- Unsupervised + Imbalanced DA
  - Matching cross-domain marginal (global) and conditional (class-wise) distributions
  - Can be simplified into TCA and JDA (if balanced UDA with atomic data distributions).
  - Example: Imbalanced DA of Office+Caltech dataset (A → C_5)
    - (a) and (b): Confusion/affinity matrix of cross-domain data using JDA and our model
    - (c) and (d): t-SNE 2D visualization using JDA and our model



Hsu et al., Unsupervised Domain Adaptation with Imbalanced Cross-Domain Data, ICCV, 2015

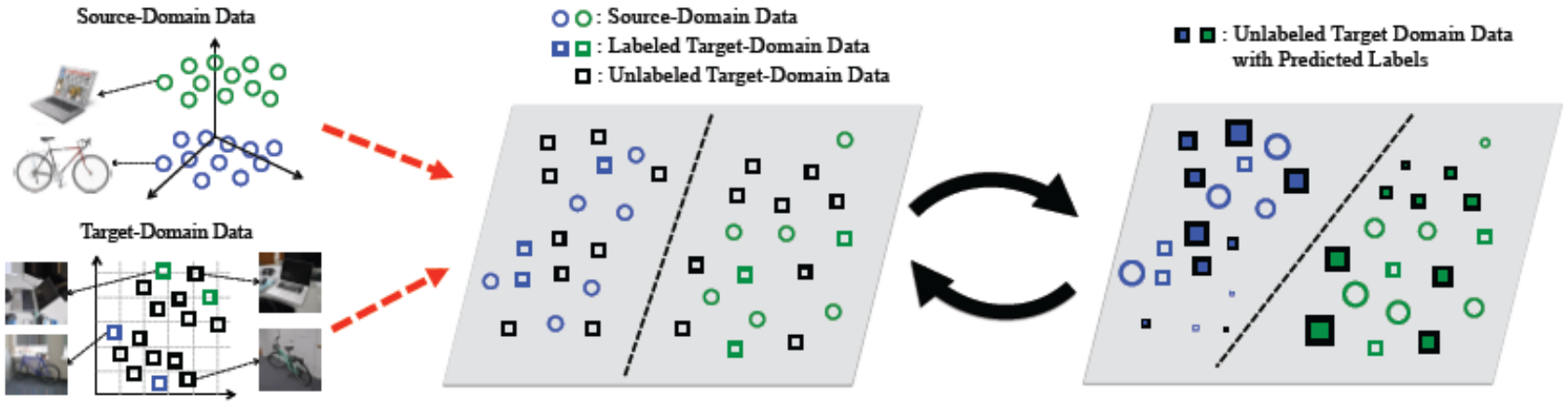
Instance + Feature Transfer

- Heterogeneous Domain Adaptation
  - Instance/landmark selection for heterogeneous cross-domain data
  - Matching cross-domain marginal and conditional distributions with selected landmarks
  - From supervised to semi-supervised heterogeneous DA

$$\min_{\mathbf{A}} E_M(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L) + E_C(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L) + \lambda \|\mathbf{A}\|^2,$$

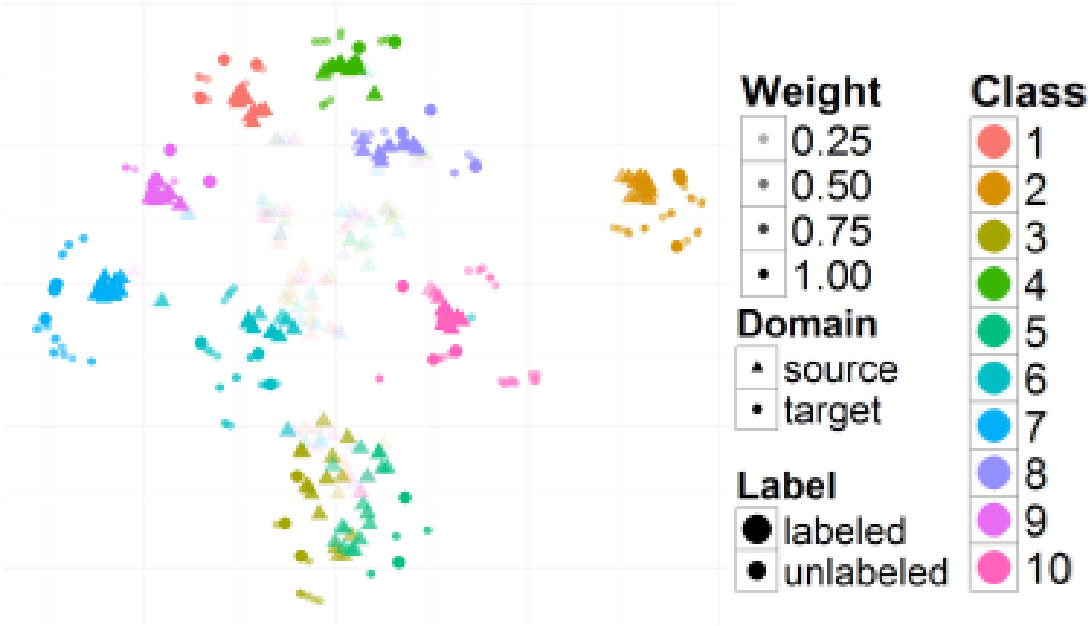
$$E_M(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L) = \left\| \frac{1}{n_S} \sum_{i=1}^{n_S} \mathbf{A}^\top \mathbf{x}_s^i - \frac{1}{n_L} \sum_{i=1}^{n_L} \widehat{\mathbf{x}}_l^i \right\|^2$$

$$E_C(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L) = \sum_{c=1}^{C} \left\| \frac{1}{n_S^c} \sum_{i=1}^{n_S^c} \mathbf{A}^\top \mathbf{x}_s^{i,c} - \frac{1}{n_L^c} \sum_{i=1}^{n_L^c} \widehat{\mathbf{x}}_l^{i,c} \right\|^2 + \frac{1}{n_S^c n_L^c} \sum_{i=1}^{n_S^c} \sum_{j=1}^{n_L^c} \left\| \mathbf{A}^\top \mathbf{x}_s^{i,c} - \widehat{\mathbf{x}}_l^{j,c} \right\|^2$$



Instance + Feature Transfer

- Heterogeneous Domain Adaptation
  - Instance/landmark selection for heterogeneous cross-domain data
  - Matching cross-domain marginal and conditional distributions with selected landmarks
  - From supervised to semi-supervised heterogeneous DA
  - Example visualization:
    Office+Caltech: Caltech (SURF) → DSLR ($DeCAF_6$)



Tsai, Yeh, and Wang, Learning Cross-Domain Landmarks for Heterogeneous Domain Adaptation, CVPR, 2016

Approaches for Transfer Learning

- Instance Transfer
  - Re-weight source-domain label instances for adaptation
- Feature Transfer
  - Derive common feat… cross-domain data
- Parameter Trans…
  - Discover sha… data
  - E.g., Domai…

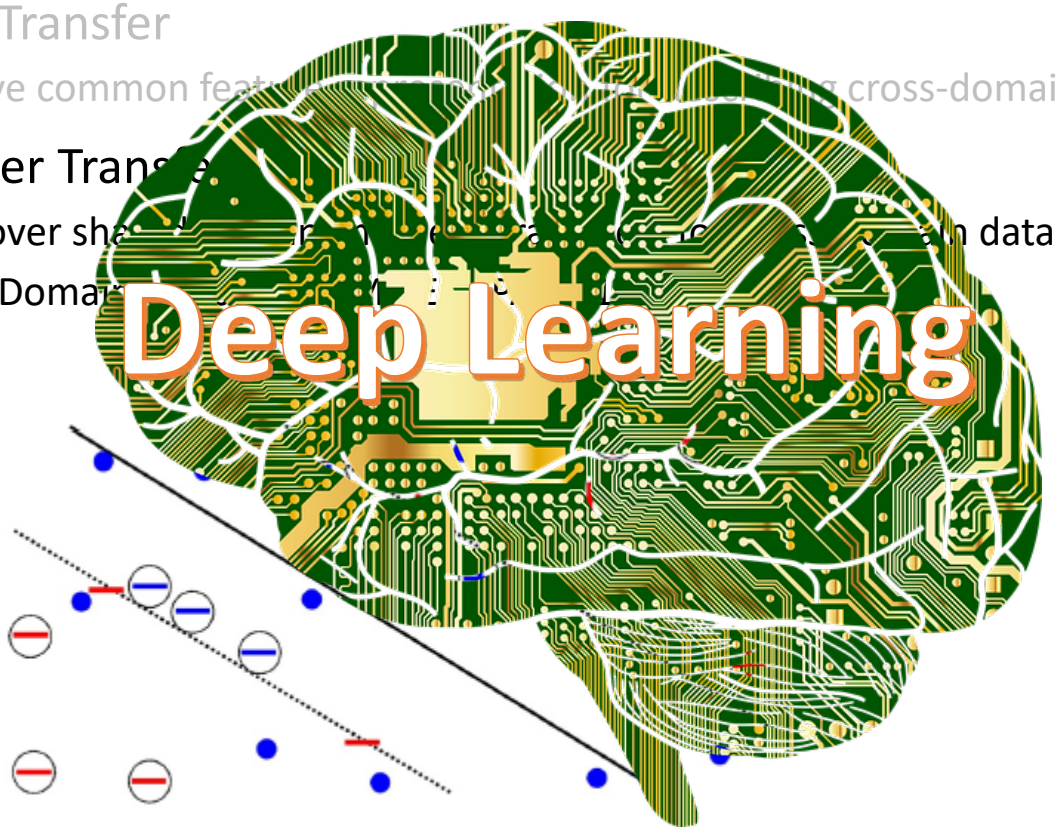
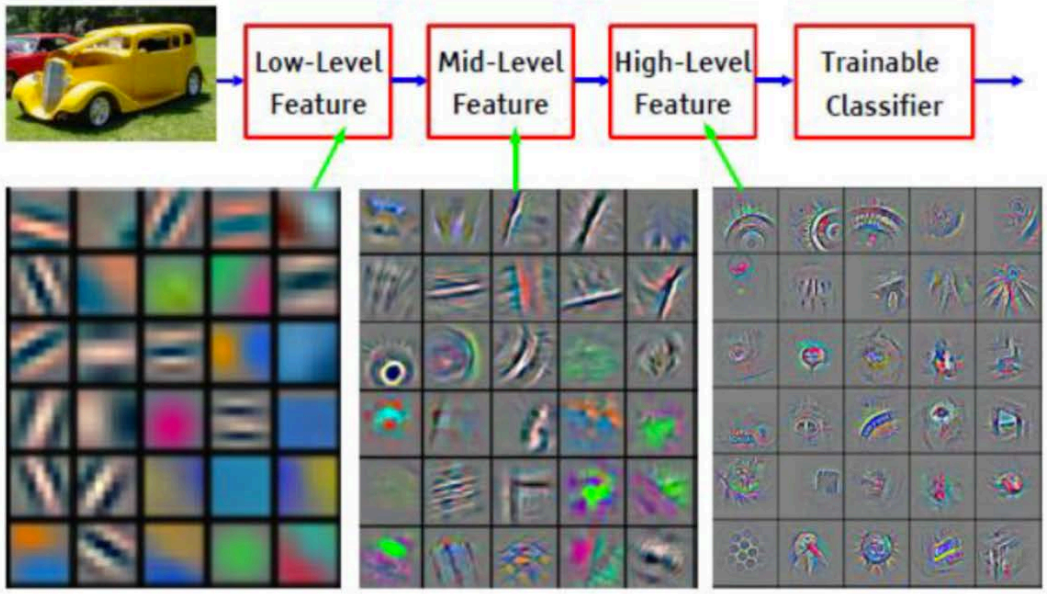


Image: Courtesy to Amaury Habrard.

Deep Feature is Sufficiently Promising.

- Convolutional Neural Networks
  - High non-linearity at different scales/layers extract features with representation power while invariant to image backgrounds, domains, etc.

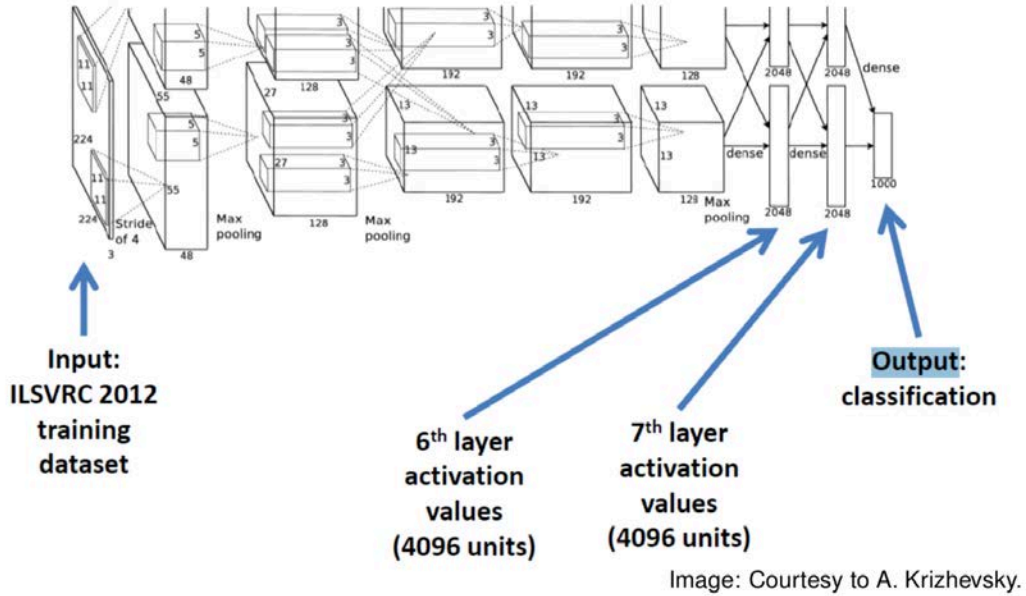


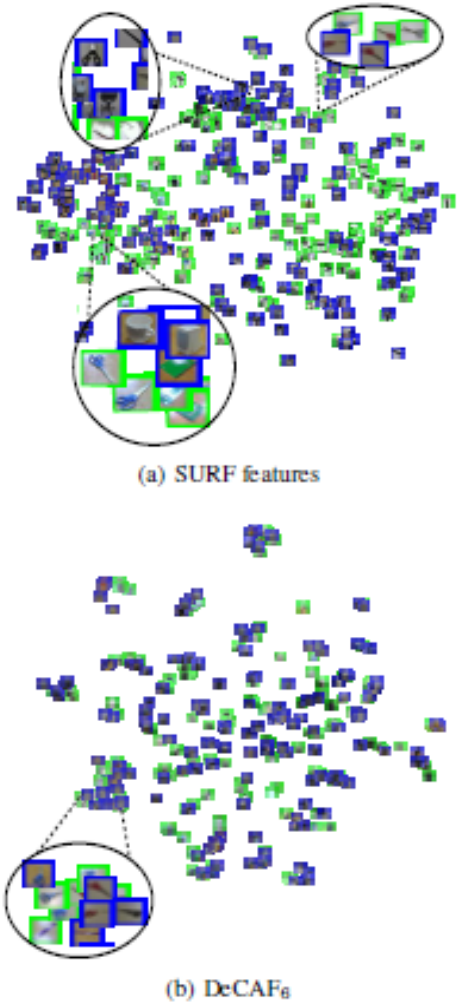
Deep Learning: More than one stage of non-linear feature transformation

Image: Courtesy to Y LeCun H.

Deep Feature is Sufficiently Promising.

- DeCAF
  - Leveraging an auxiliary large dataset to train CNN.
  - The resulting features exhibit sufficient representation ability.
  - Supporting results on Office+Caltech datasets, etc.

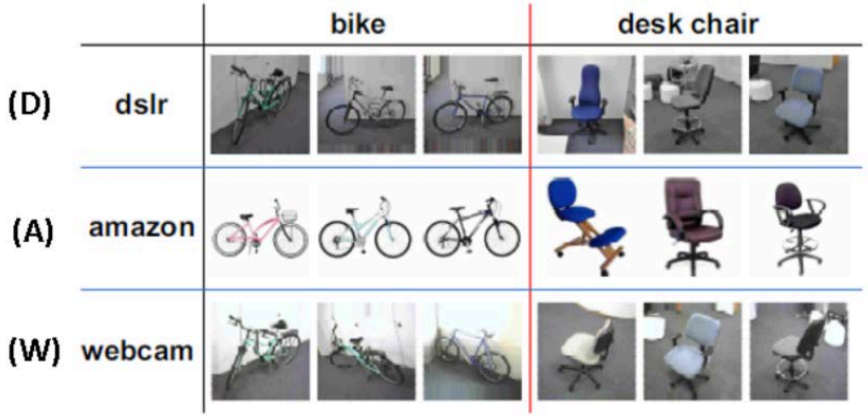




Donahue et al., DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition, ICML 2014

Deep Feature is Sufficiently Promising.

- DeCAF
  - Leveraging an auxiliary large dataset to train CNN.
  - The resulting features exhibit sufficient representation ability.
  - Supporting results on Office+Caltech datasets, etc.



Feature	SURF											$Decaf_6$				
data	Raw	SA	SDA	GFK	TCA	JDA	TJM	SCA	JGSA primal	JGSA linear	JGSA RBF	JDA	OTGL	JGSA primal	JGSA linear	JGSA RBF
A→D	35.67	33.76	33.76	40.13	33.76	39.49	45.22	39.49	**47.13**	45.86	45.22	81.53	85.00	**88.54**	85.35	84.71
A→W	31.19	33.22	30.85	36.95	36.27	37.97	42.03	34.92	45.76	**49.49**	45.08	80.68	83.05	81.02	**84.75**	80.00
D→A	28.29	**39.87**	38.73	28.71	31.00	33.09	32.78	31.63	38.00	36.01	38.73	91.96	**92.31**	91.96	92.28	91.96
D→W	83.73	76.95	76.95	80.34	86.10	89.49	85.42	84.41	91.86	91.86	**93.22**	99.32	96.29	**99.66**	98.64	98.64
W→A	31.63	39.25	39.25	27.56	28.91	32.78	29.96	29.96	39.87	**41.02**	40.81	90.71	90.62	90.71	**91.44**	91.34
W→D	84.71	75.16	75.80	85.35	89.17	89.17	89.17	87.26	**90.45**	**90.45**	88.54	**100**	96.25	**100**	**100**	**100**

Donahue et al., DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition, ICML 2014

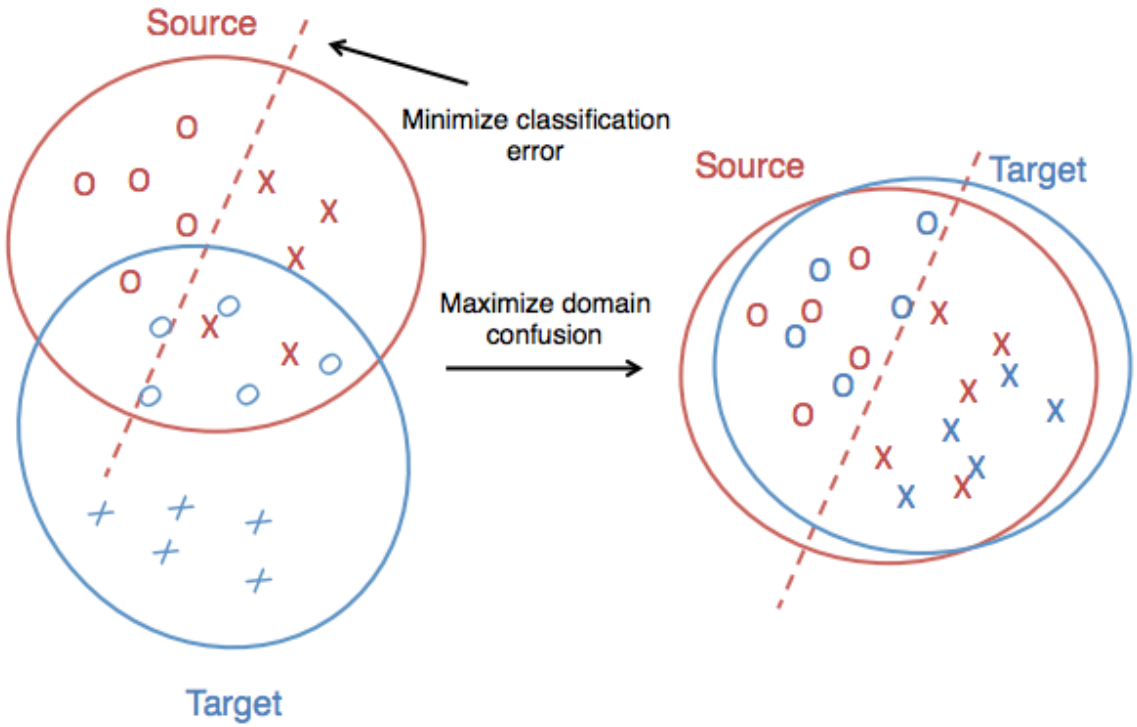
Recent Deep Learning Methods for TL

- Deep Domain Confusion (DDC)
- Domain-Adversarial Training of Neural Networks (DANN)
- Adversarial Discriminative Domain Adaptation (ADDA)
- Domain Separation Network (DSN)
- Unsupervised Pixel-Level Domain Adaptation with Generative Adversarial Networks (PixelDA)
- No More Discrimination: Cross City Adaptation of Road Scene Segmenters

	Shared weights	Adaptation loss	Generative model
DDC	✓	MMD	✗
DANN	✓	Adversarial	✗
ADDA	✗	Adversarial	✗
DSN	Partially shared	MMD/Adversarial	✗
PixelDA	✗	Adversarial	✓

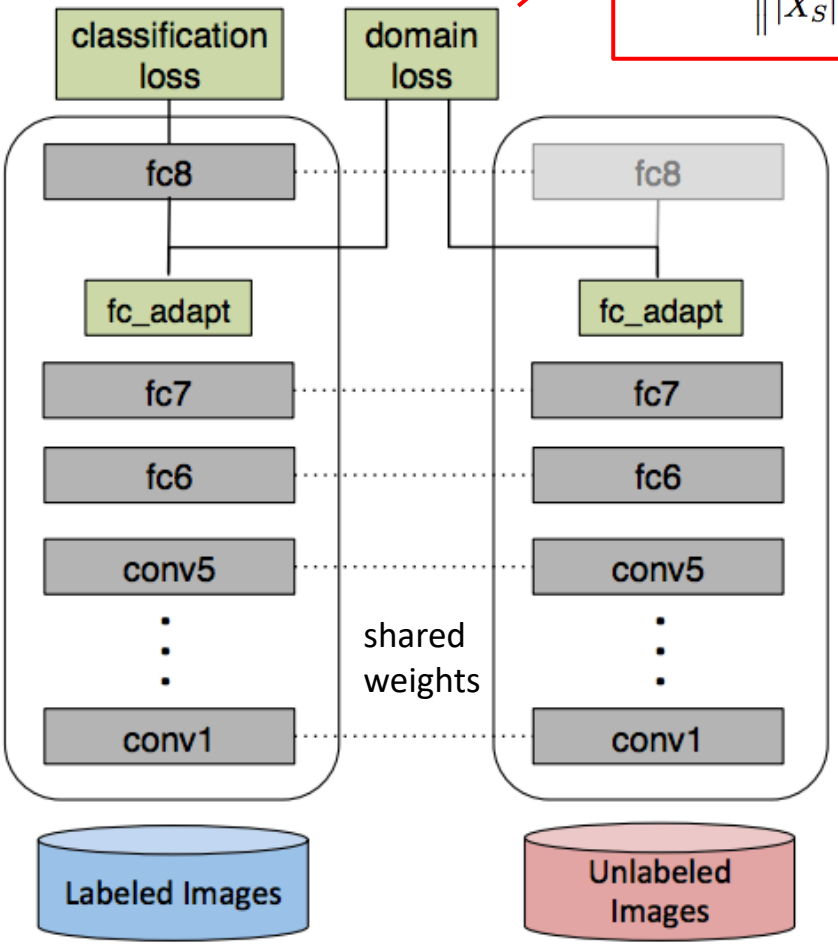
Deep Domain Confusion (DDC)

- Deep Domain Confusion: Maximizing for Domain Invariance
  - Tzeng et al., arXiv: 1412.3474, 2014



Deep Domain Confusion (DDC)

$$\mathrm{MMD}(X_S, X_T) = \left\| \frac{1}{|X_S|} \sum_{x_s \in X_S} \phi(x_s) - \frac{1}{|X_T|} \sum_{x_t \in X_T} \phi(x_t) \right\|$$

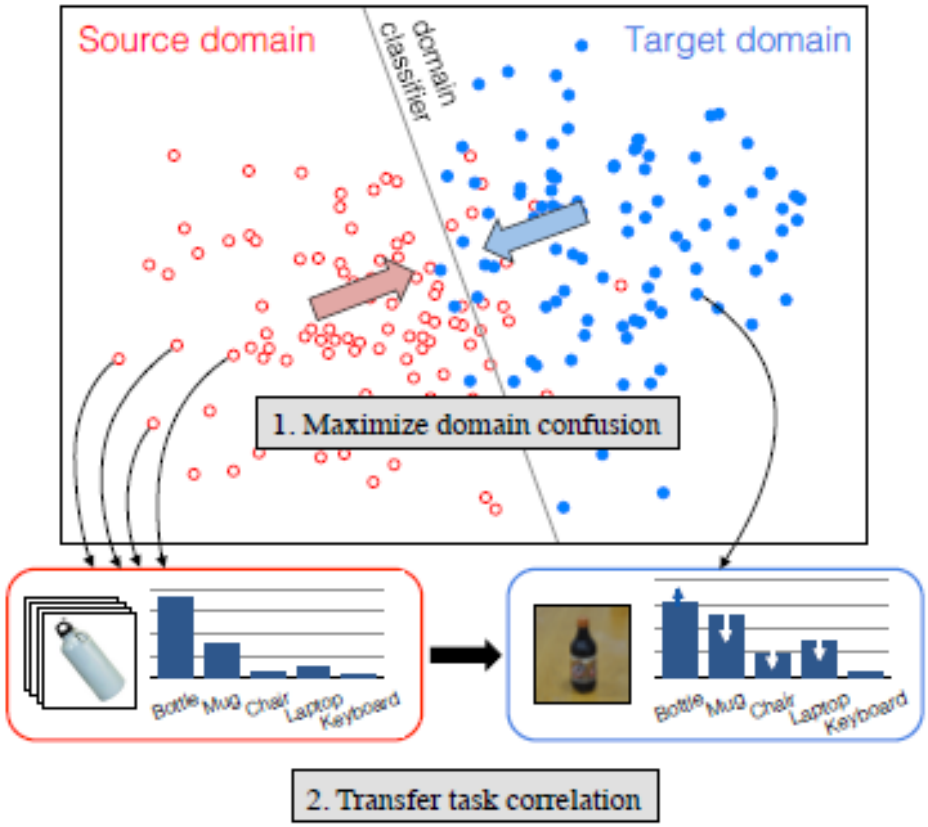


✓Minimize classification loss:

$$\mathcal{L} = \mathcal{L}_C(X_L, y) + \lambda \, \mathrm{MMD}^2(X_S, X_T)$$

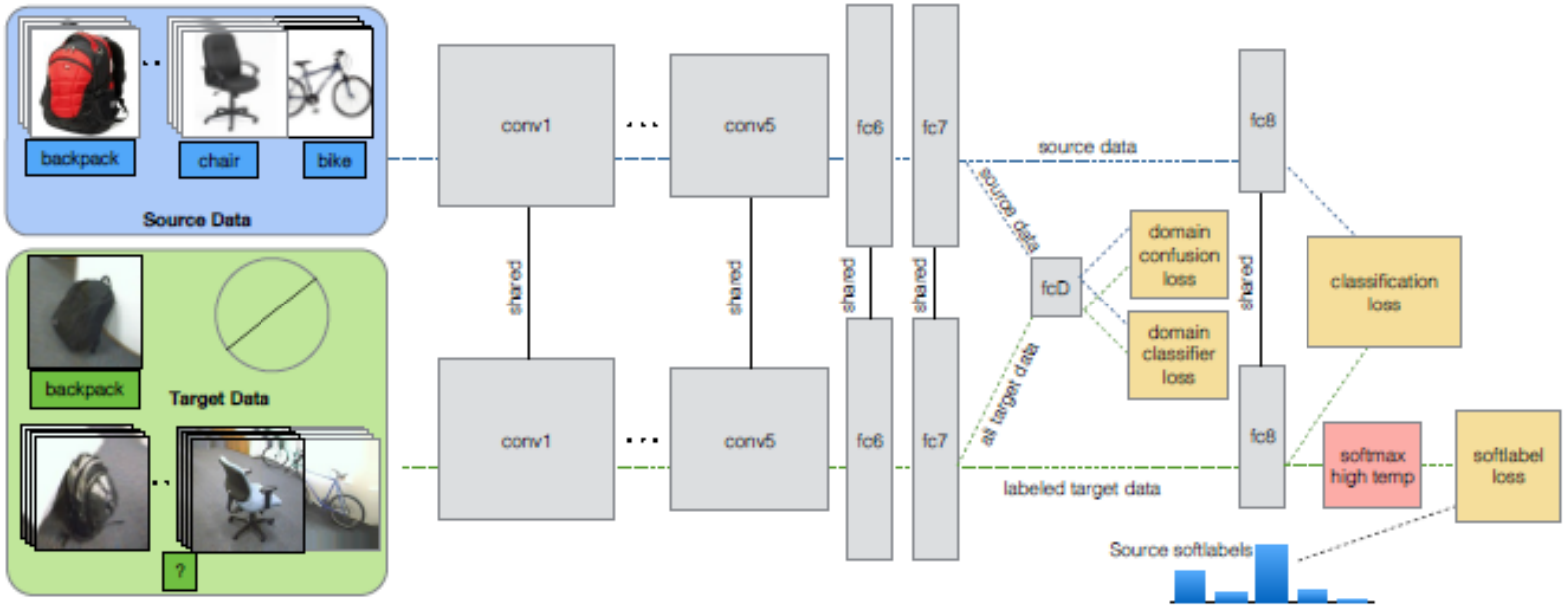
Deep Domain Confusion (DDC)

- Simultaneous Deep Transfer Across Domains and Tasks
  - Tzeng et al., ICCV, 2015



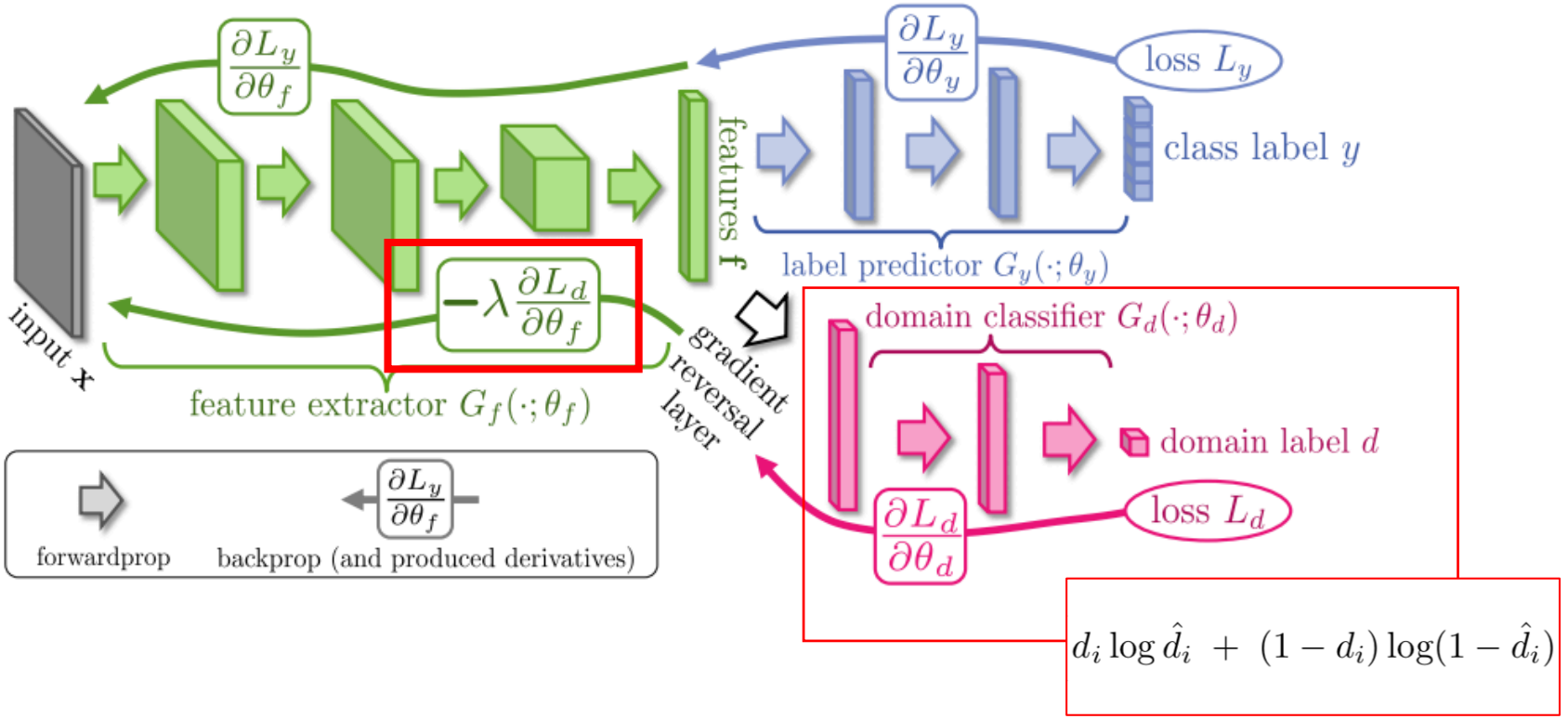
Deep Domain Confusion (DDC)

- Simultaneous Deep Transfer Across Domains and Tasks
  - Tzeng et al., ICCV, 2015
  - Soft label loss is additionally introduced.



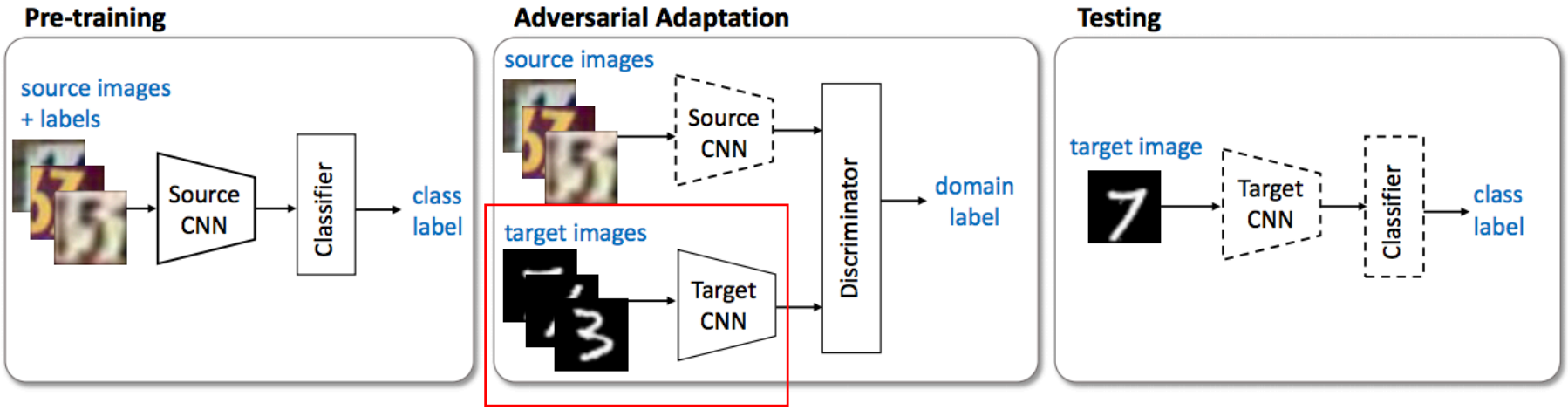
Domain Confusion by Domain-Adversarial Training

- Domain-Adversarial Training of Neural Networks (DANN)
  - Y. Ganin et al., ICML 2015
  - Maximize domain confusion = maximize domain classification loss
  - Minimize source-domain data classification loss



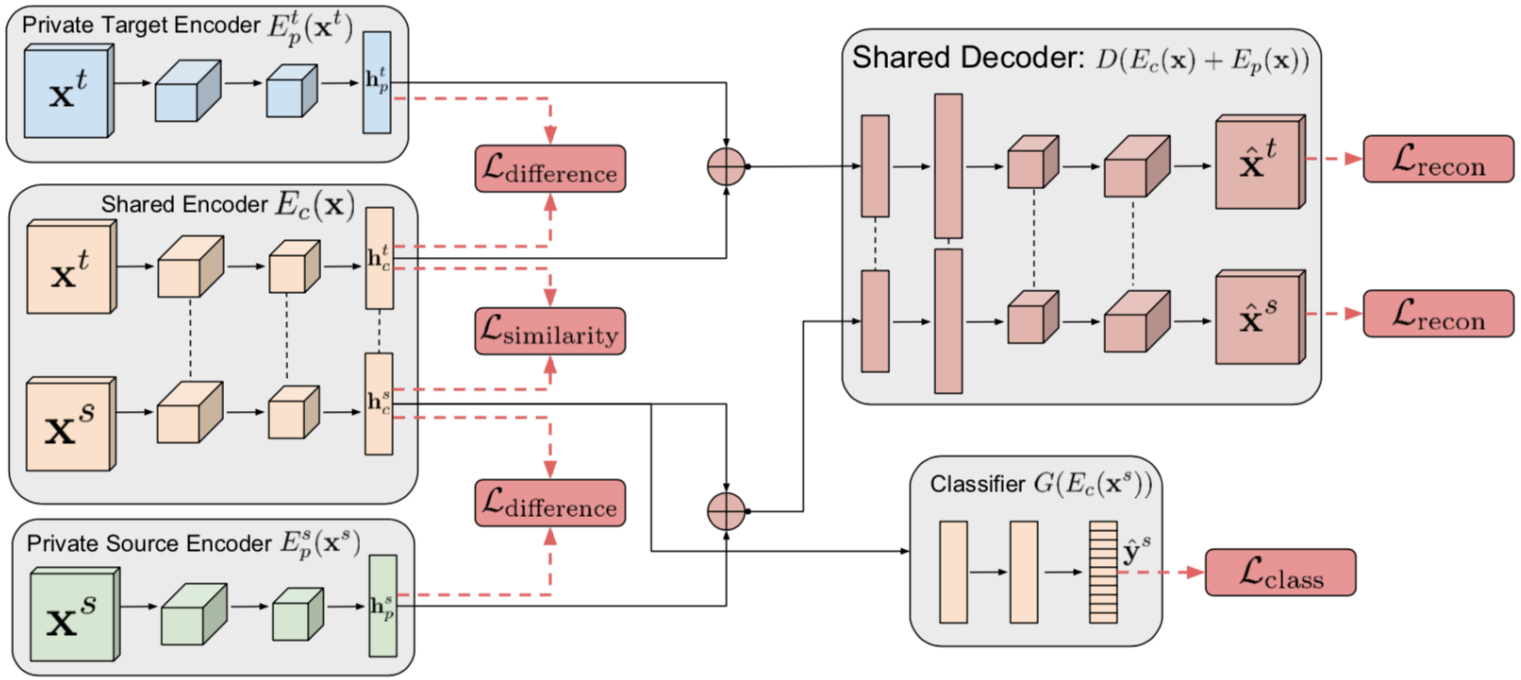
Domain Confusion by Domain-Adversarial Training

- Adversarial Discriminative Domain Adaptation
  - Tzeng et al., CVPR 2017
  - Maximize domain confusion = maximize domain classification loss
  - Minimize source-domain data classification loss
  - Compared to DANN, a distinct decoder for the target domain is considered.



Beyond Domain Confusion

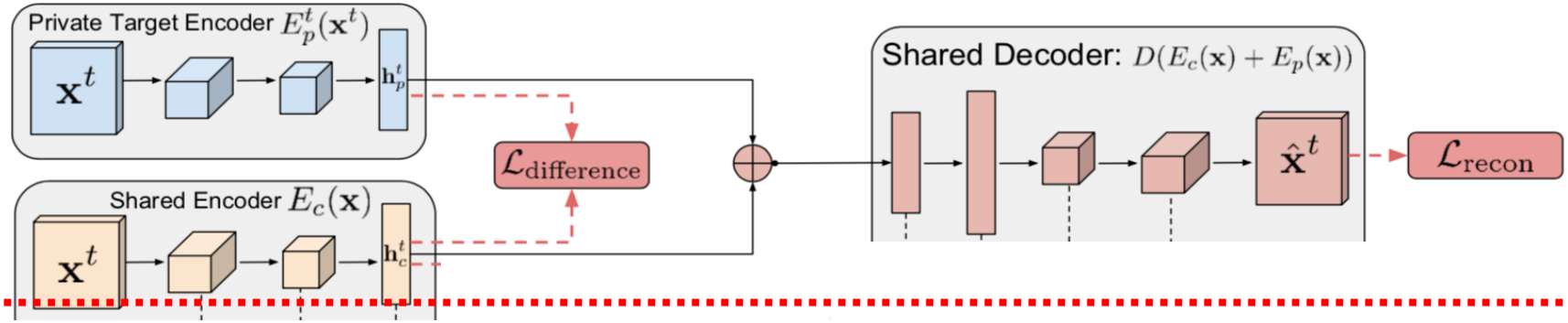
- Domain Separation Network
  - Bousmalis et al., NIPS 2016
  - Separate encoders for domain-invariant and domain-specific features



Beyond Domain Confusion

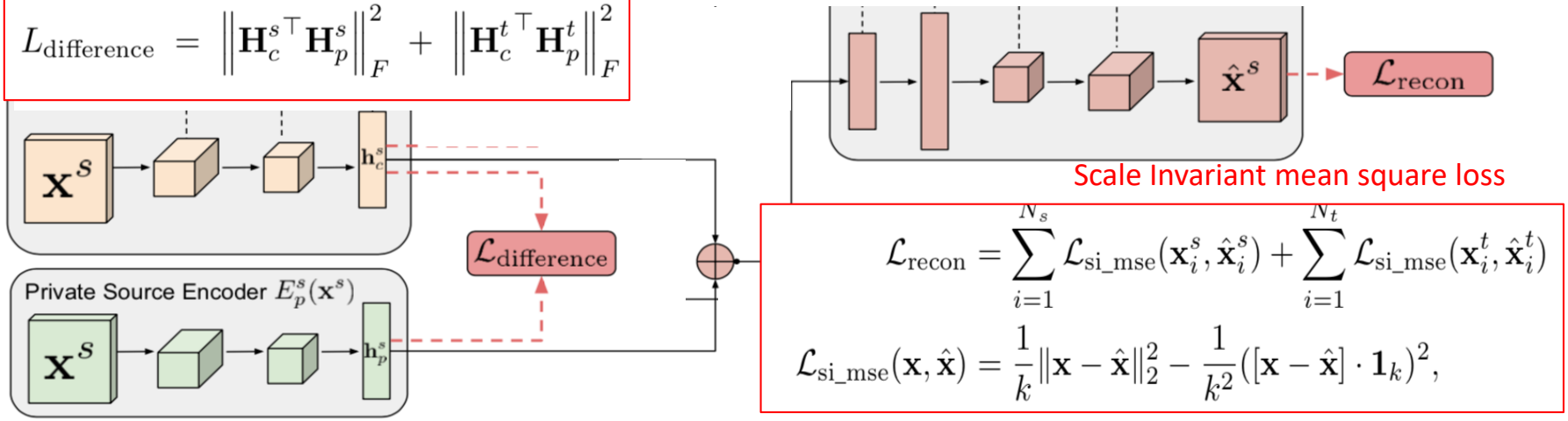
- Domain Separation Network, NIPS 2016

✓Auto-encoder structure preserves visual information in each domain



Orthogonality difference loss

$$L_{\text{difference}} = \left\| \mathbf{H}_c^{s\top} \mathbf{H}_p^s \right\|_F^2 + \left\| \mathbf{H}_c^{t\top} \mathbf{H}_p^t \right\|_F^2$$



Scale Invariant mean square loss

$$\mathcal{L}_{\text{recon}} = \sum_{i=1}^{N_s} \mathcal{L}_{\text{si\_mse}}(\mathbf{x}_i^s, \hat{\mathbf{x}}_i^s) + \sum_{i=1}^{N_t} \mathcal{L}_{\text{si\_mse}}(\mathbf{x}_i^t, \hat{\mathbf{x}}_i^t)$$

$$\mathcal{L}_{\text{si\_mse}}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{k} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 - \frac{1}{k^2} ([\mathbf{x} - \hat{\mathbf{x}}] \cdot \mathbf{1}_k)^2,$$

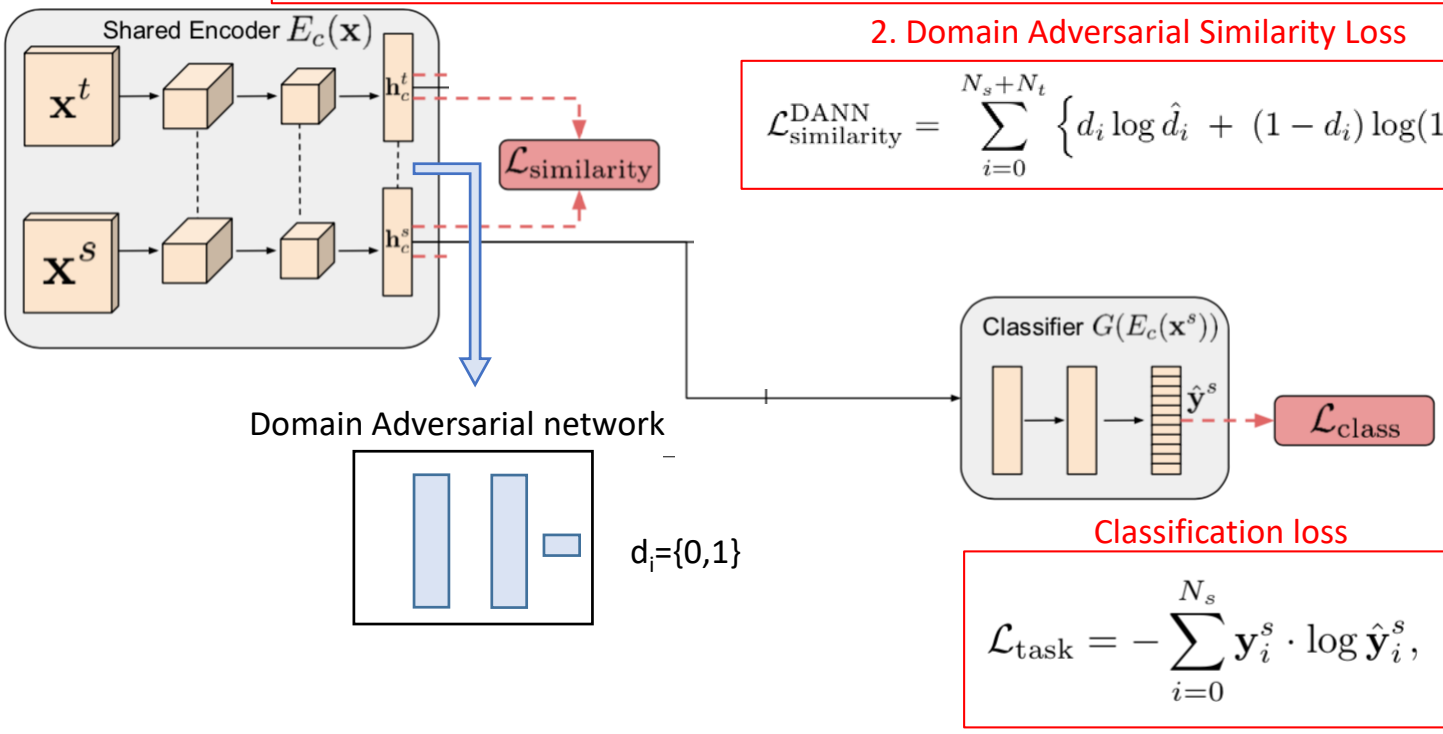
Beyond Domain Confusion

- Domain Separation Network, NIPS 2016

✓Share encoder captures domain invariant information for classification

1. Maximum Mean Discrepancy (MMD) Similarity Loss

$$\mathcal{L}_{\text{similarity}}^{\text{MMD}} = \frac{1}{(N^s)^2}\sum_{i,j=0}^{N^s}\kappa(\mathbf{h}_{ci}^s,\mathbf{h}_{cj}^s) - \frac{2}{N^sN^t}\sum_{i,j=0}^{N^s,N^t}\kappa(\mathbf{h}_{ci}^s,\mathbf{h}_{cj}^t) + \frac{1}{(N^t)^2}\sum_{i,j=0}^{N^t}\kappa(\mathbf{h}_{ci}^t,\mathbf{h}_{cj}^t)$$



2. Domain Adversarial Similarity Loss

$$\mathcal{L}_{\text{similarity}}^{\text{DANN}} = \sum_{i=0}^{N_s+N_t}\left\{d_i\log\hat{d}_i + (1-d_i)\log(1-\hat{d}_i)\right\}.$$

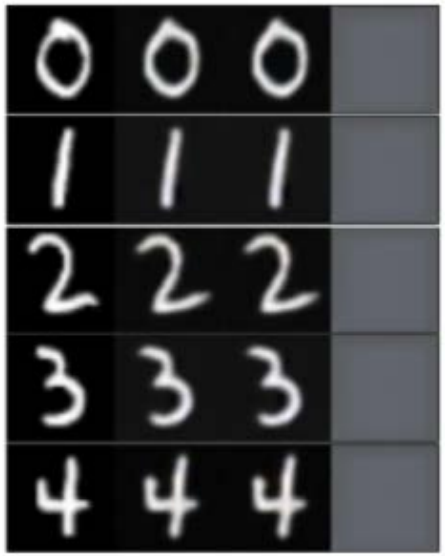
Classification loss

$$\mathcal{L}_{\text{task}} = -\sum_{i=0}^{N_s}\mathbf{y}_i^s\cdot\log\hat{\mathbf{y}}_i^s,$$

Beyond Domain Confusion

- Domain Separation Network, NIPS 2016
  - Example results

X_t Original image	$D(E_c(x_t)+E_p(x_t))$ Reconstruct private + share	$D(E_c(x_t))$ Reconstruct Share only	$D(E_p(x_t))$ Reconstruct Private only

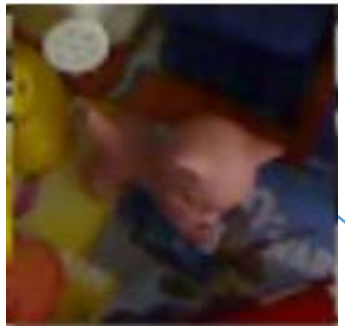


(a) MNIST (source)

(b) MNIST-M (target)

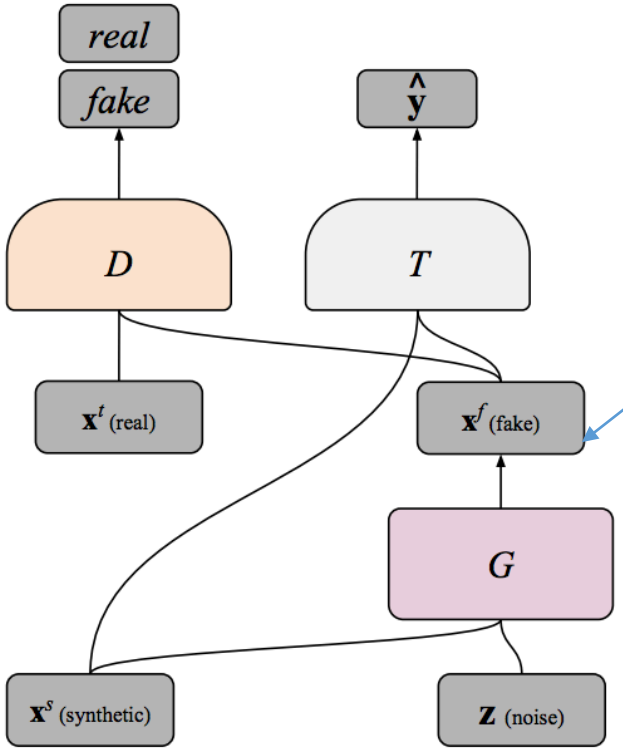
Generative Models for Transfer Learning

- Unsupervised Pixel-Level Domain Adaptation with Generative Adversarial Networks
  - Bousmalis et al., CVPR 2017
  - Advance generative models for domain adaptation tasks



Real data as target domain

Synthesized data as source domain



Generated faked data

Generative Models for Transfer Learning

- Unsupervised Pixel-Level Domain Adaptation with Generative Adversarial Networks
  - Bousmalis et al., CVPR 2017
  - Advance generative models for domain adaptation tasks

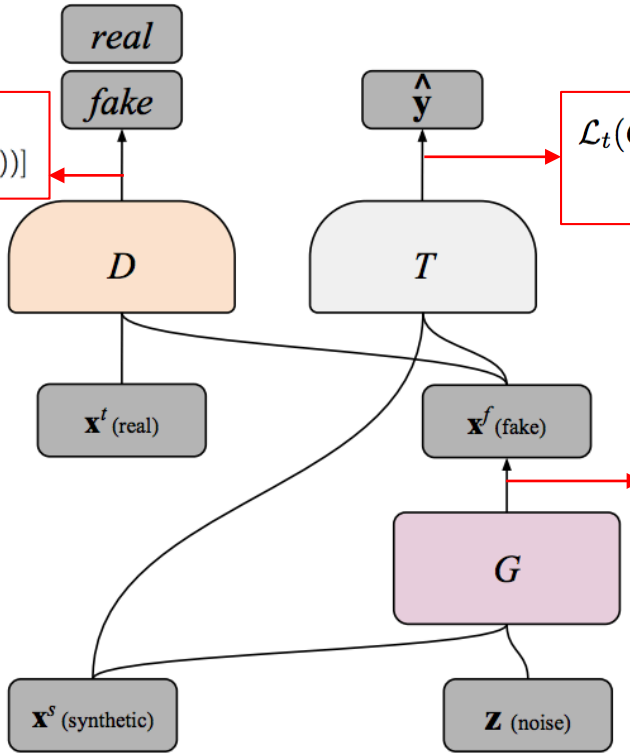
✓The classifier is trained on (synthetic) source-domain data and fake (real) target-domain data.

Domain Adversarial Loss

$$\mathcal{L}_d(D,G) = \mathbb{E}_{\mathbf{x}^t}[\log D(\mathbf{x}^t;\boldsymbol{\theta}_D)] + \mathbb{E}_{\mathbf{x}^s,\mathbf{z}}[\log(1 - D(G(\mathbf{x}^s,\mathbf{z};\boldsymbol{\theta}_G);\boldsymbol{\theta}_D))]$$

Classification Cross Entropy

$$\mathcal{L}_t(G,T) = \mathbb{E}_{\mathbf{x}^s,\mathbf{y}^s,\mathbf{z}}\big[-\mathbf{y}^{s\top} \log T\left(G(\mathbf{x}^s,\mathbf{z};\boldsymbol{\theta}_G);\boldsymbol{\theta}_T\right) - \mathbf{y}^{s\top} \log T(\mathbf{x}^s);\boldsymbol{\theta}_T\big] \quad (3)$$



Masked Pairwise Mean Squared Error

$$\mathcal{L}_c(G) = \mathbb{E}_{\mathbf{x}^s,\mathbf{z}}\Big[\frac{1}{k}\left\|(\mathbf{x}^s - G(\mathbf{x}^s,\mathbf{z};\boldsymbol{\theta}_G)) \circ \mathbf{m}\right\|_2^2 - \frac{1}{k^2}\left((\mathbf{x}^s - G(\mathbf{x}^s,\mathbf{z};\boldsymbol{\theta}_G))^\top \mathbf{m}\right)^2\Big]$$

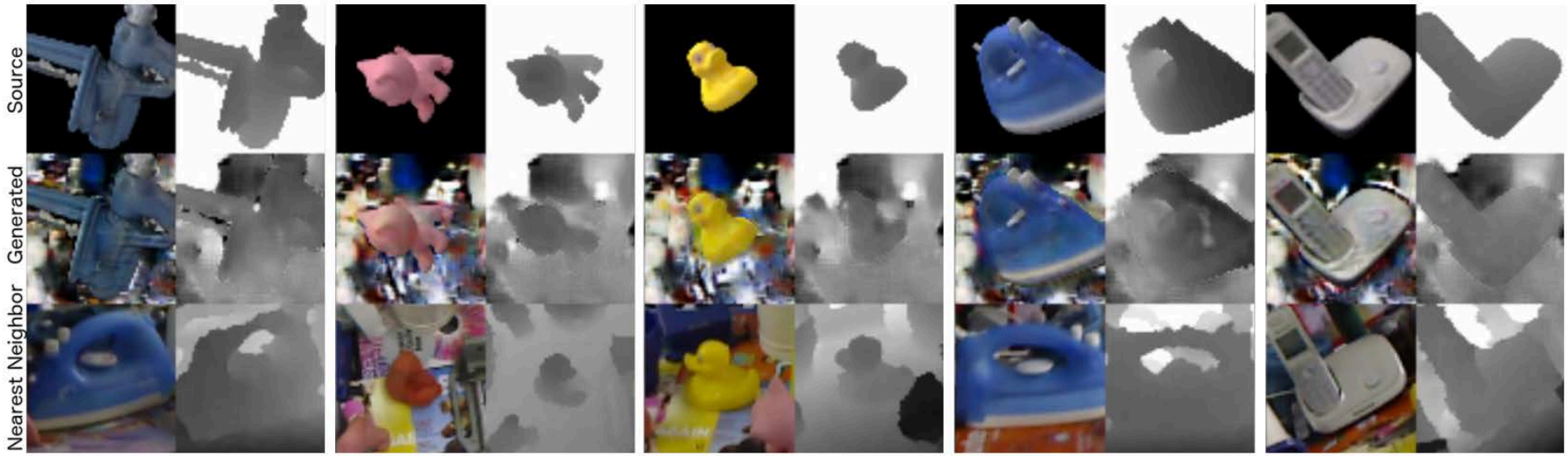
Mask

Fake image

Nearest real image

Generative Models for Transfer Learning

- Unsupervised Pixel-Level Domain Adaptation with Generative Adversarial Networks
  - Bousmalis et al., CVPR 2017
  - Advance generative models for domain adaptation tasks

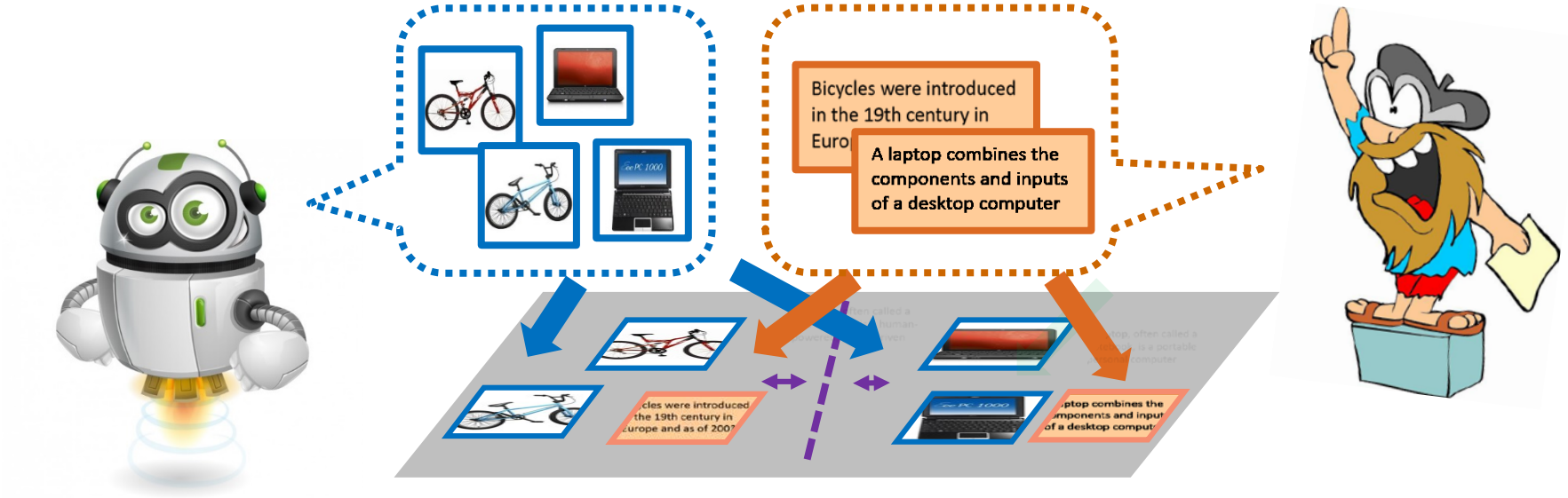


Deep Transfer Learning for *Cross-Domain Data Classification*

- Heterogeneous Domain Adaptation
  - Learning from source & target-domain data

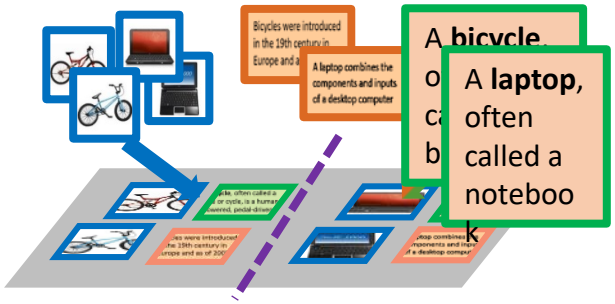
    described by different types of features

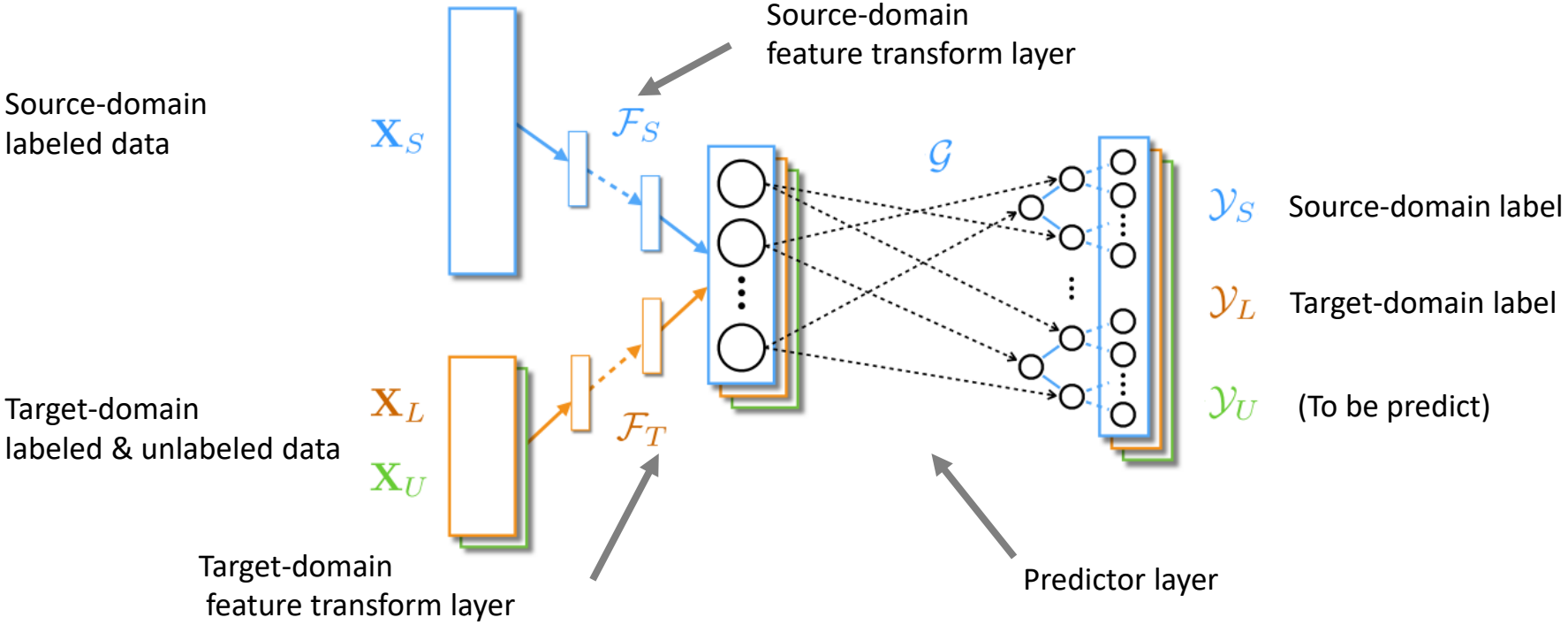
    (e.g., text/speech-to-image classification, cross-sensor data analysis)



Deep Transfer Learning for *Cross-Domain Data Classification*

- Heterogeneous Domain Adaptation
  - Learning from source & target-domain data described by different types of features (e.g., text/speech-to-image classification, cross-sensor data analysis)





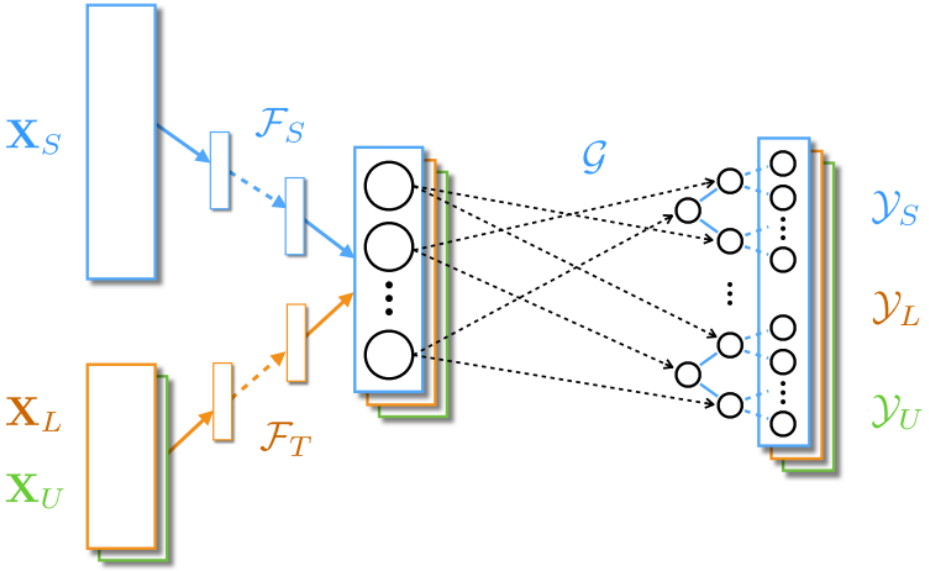
Deep Transfer Learning for *Cross-Domain Data Classification* (cont'd)

- Our Proposed Method: ***Transfer Neural Trees (TNT)*** [ECCV'16]
- *Highlights*:
  - Joint learning of cross-domain mapping F_S/F_T & cl. layer G (deep neural decision forest)
  - Propose stochastic pruning for G to avoid overfitting source-domain labeled data
  - Unique embedding loss for learning target-domain data in a *semi-supervised* setting

Source-domain labeled data

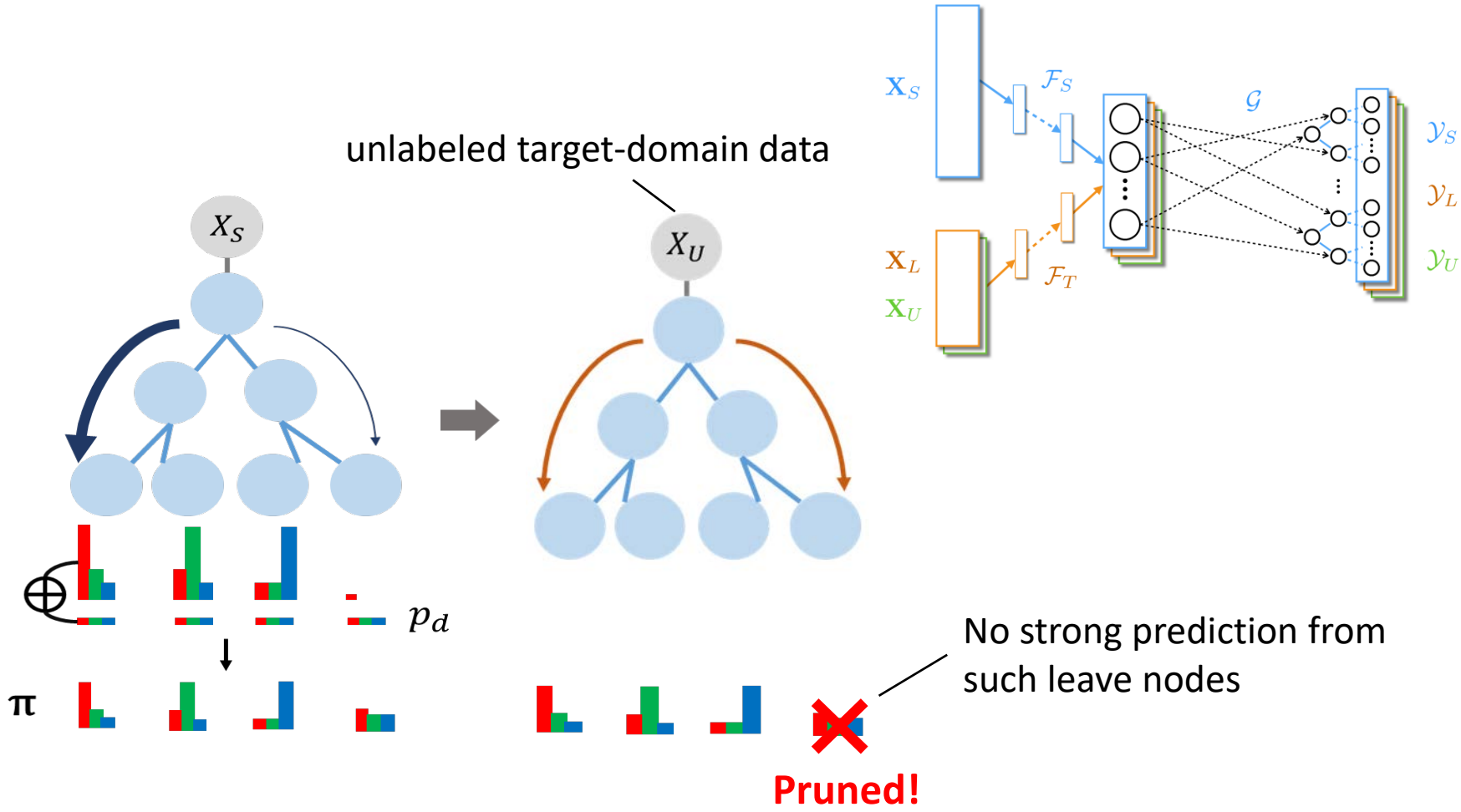
Target-domain labeled data

Target-domain unlabeled data



Wang et al., "Transfer Neural Trees for Heterogeneous Domain Adaptation," ECCV, 2016.

Preventing Overfitting in Prediction Layer G



Wang et al., "Transfer Neural Trees for Heterogeneous Domain Adaptation," ECCV, 2016.

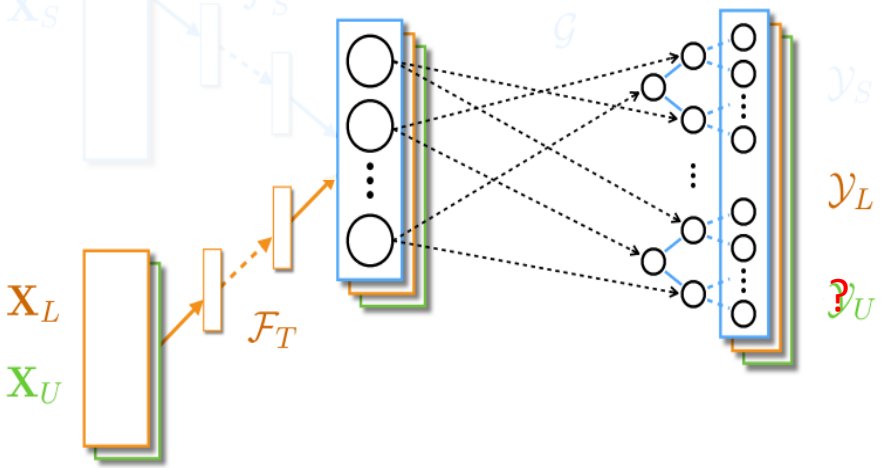
We Still Need…

Semi-Supervised Learning for Target-Domain Data

Learning of F_T

with Prediction loss L_p of target-domain labeled data
and Embedding loss L_e of target-domain labeled & unlabeled data

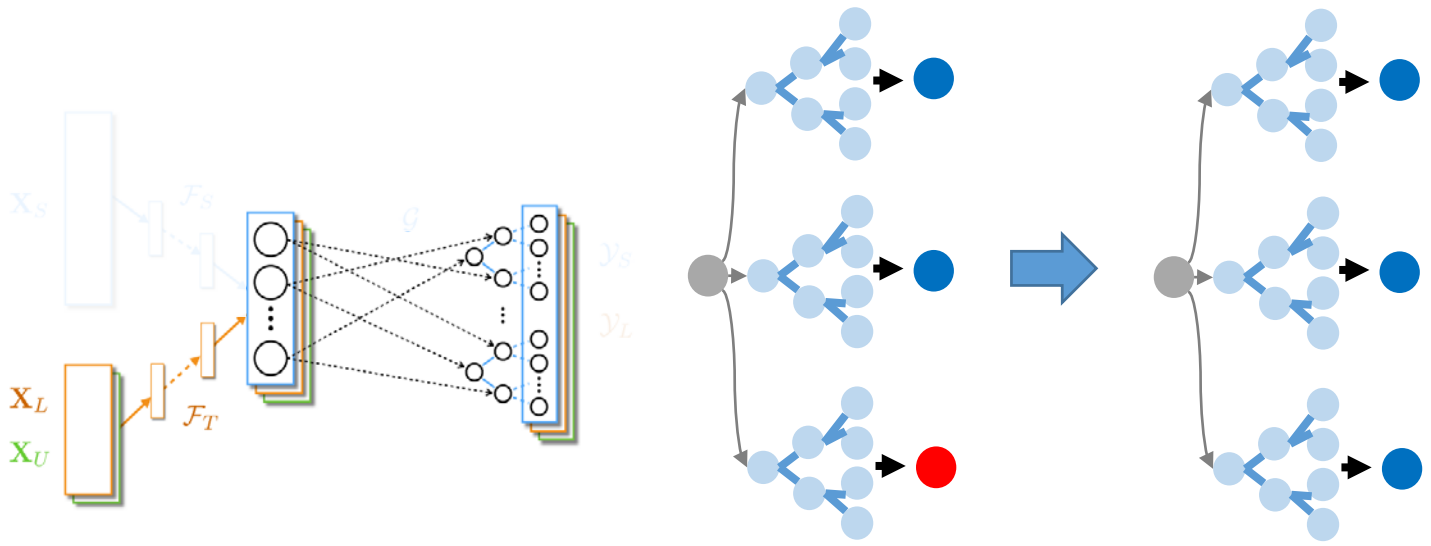
$$\min_{\mathcal{F}_T} \sum_{(\mathbf{x},y)\in D_L} L_p(\Theta, \boldsymbol{\pi}; \mathbf{x}, \mathrm{y}) + \lambda \sum_{\tilde{y}} \sum_{\mathbf{x}\in\{\mathbf{X}_L,\mathbf{X}_U\}} L_e(\Theta, \boldsymbol{\pi}; \mathbf{x}, \tilde{y})$$



Wang et al., "Transfer Neural Trees for Heterogeneous Domain Adaptation," ECCV, 2016.

Embedding Loss L_e for TNT

- Goal
  - Improve similarity between mapped X_L & X_U of same class
  - However, no ground truth Labels for X_U
- Solution
  - enforce prediction consistency to preserve structure consistency



Wang et al., "Transfer Neural Trees for Heterogeneous Domain Adaptation," ECCV, 2016.

Example Results for *Cross-Domain Data Classification*

- Cross-Domain Object Recognition
  - Datasets: Office + Caltech256 [Saenko et al., ECCV'10, G. Griffin et al., 2007]

- Comparisons

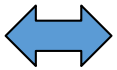
D_S -> D_T Decaf -> Surf	SVM_t	NN_t	MMDT	HFA	SHFR	SCP	TNT
C -> A	45.37	45.80	45.69	46.44	44.61	41.59	**51.62**
W -> A			46.23	46.98	43.86	44.50	**50.39**
A -> C	37.15	35.02	35.77	36.32	33.39	35.04	**37.53**
W -> C			36.05	36.41	33.21	35.96	**39.52**
A -> W	61.51	61.06	61.13	61.89	54.34	58.87	**64.38**
C -> W			60.76	62.26	54.34	51.32	**66.64**

Wang et al., "Transfer Neural Trees for Heterogeneous Domain Adaptation," ECCV, 2016.

Deep Transfer Learning for *Multi-Label Classification*

- Multi-label classification
  - Predicting multiple labels without observing annotated ground truth info
  - Learning across image and label-domain data + exploit label co-occurrences

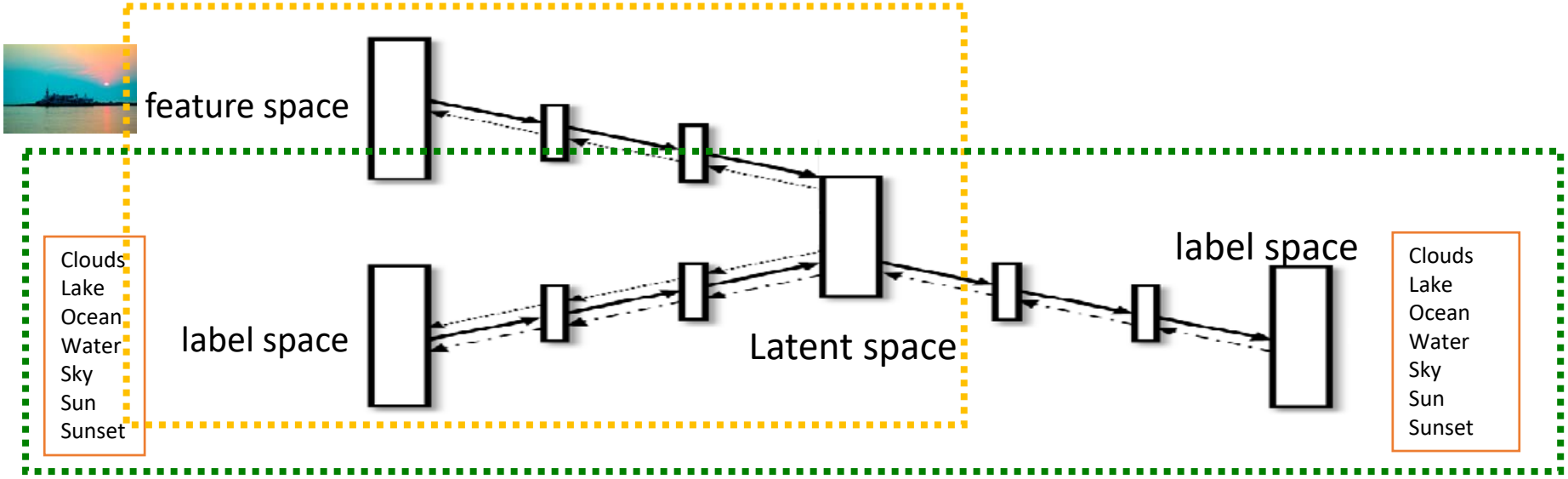
Image



Labels:
Person
Table
Sofa
Chair
TV
Lights
Carpet
…

Deep Transfer Learning for *Multi-Label Classification*

- Our Proposed Method: ***Canonical Correlated AutoEncoder (C2AE)*** [AAAI'17]
- *Highlights*:
  - Unique integration of autoencoder & deep canonical correlation analysis (DCCA)
  - Autoencoder in C2AE: label embedding + label recovery + label co-occurrence
  - DCCA in C2AE: joint feature & label embedding
  - Can handle missing labels during learning



Wang et al., Learning Deep Latent Spaces for Multi-Label Classification, AAAI 2017

Deep Transfer Learning for *Multi-Label Classification*

- Our Proposed Method: ***Canonical Correlated AutoEncoder (C2AE)*** [AAAI'17]
- *Highlights*:
  - Unique integration of autoencoder & deep canonical correlation analysis (DCCA)

Wang et al., Learning Deep Latent Spaces for Multi-Label Classification, AAAI 2017

Semantic Segmentation Across Cities

- No More Discrimination: Cross City Adaptation of Road Scene Segmenters
  - Chen et al., ICCV 2017
  - Weakly supervised DA for semantic segmentation

?

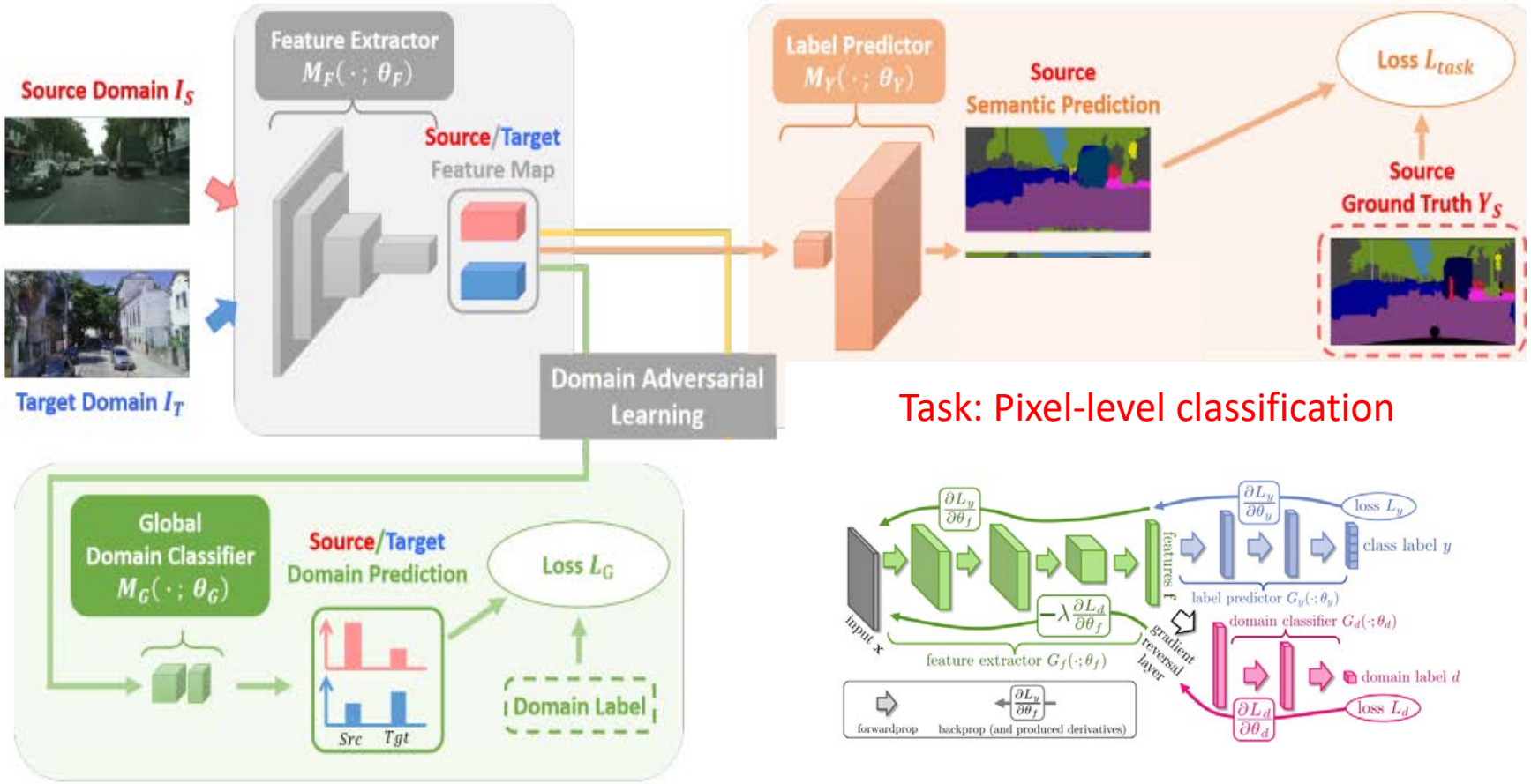
?

?

?

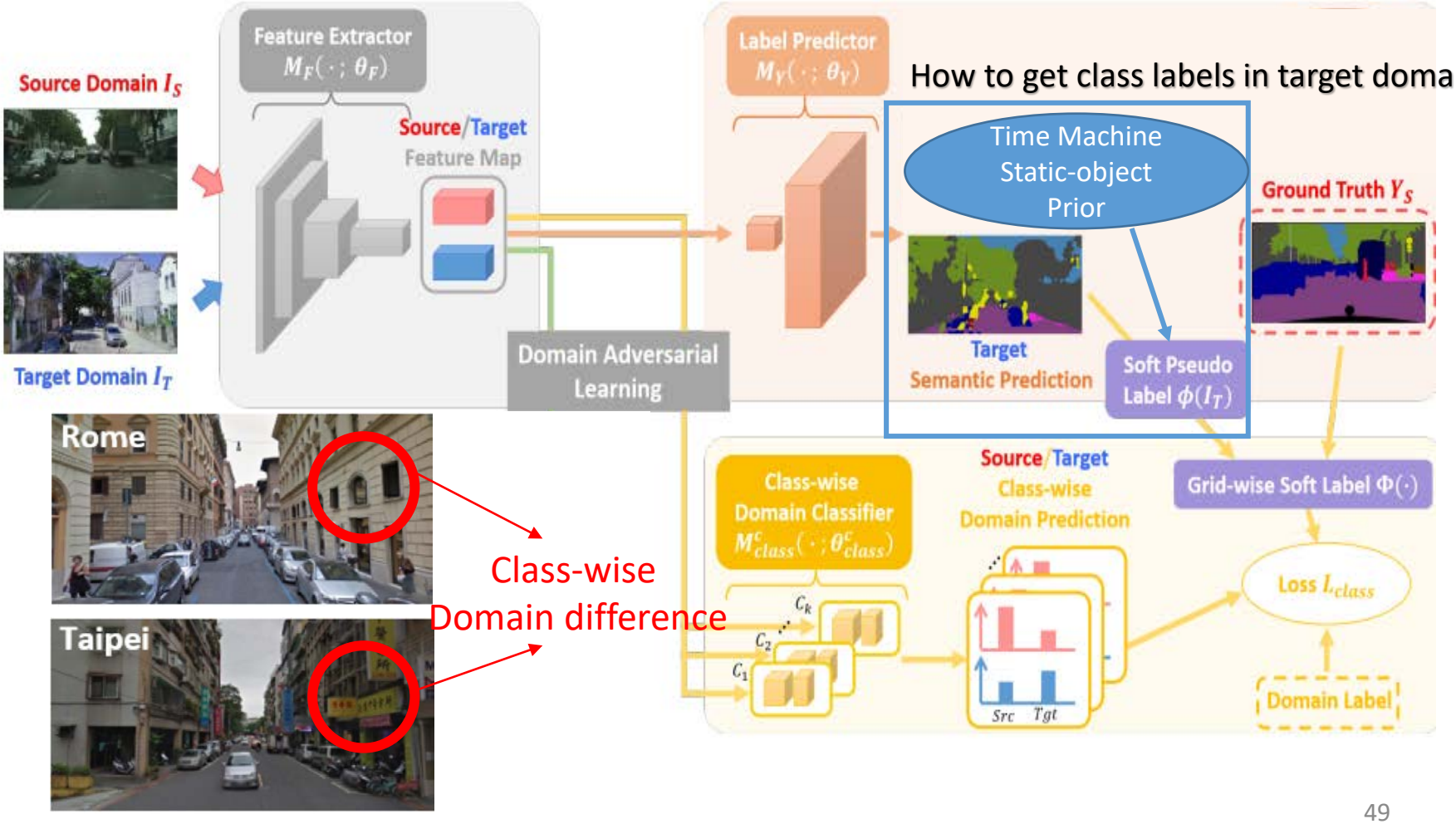
Semantic Segmentation Across Cities

- No More Discrimination: Cross City Adaptation of Road Scene Segmenters
  - Chen et al., ICCV 2017
  - Weakly supervised DA for semantic segmentation



Semantic Segmentation Across Cities

- No More Discrimination: Cross City Adaptation of Road Scene Segmenters



Semantic Segmentation Across Cities

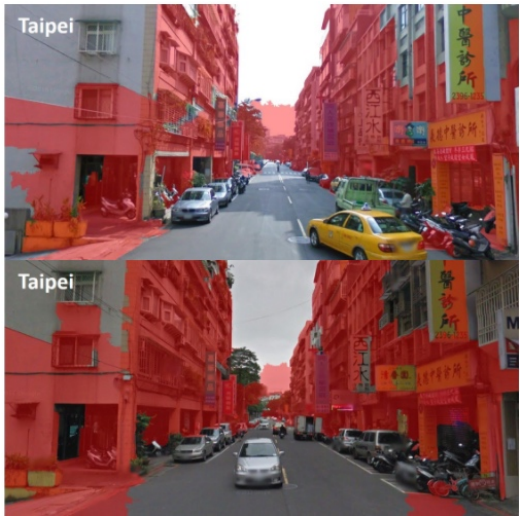
- No More Discrimination: Cross City Adaptation of Road Scene Segmenters
  - Chen et al., ICCV 2017
  - Weakly supervised DA for semantic segmentation
  - Static-object prior from Google Map Time Machine features

2015

2016

Dense Match

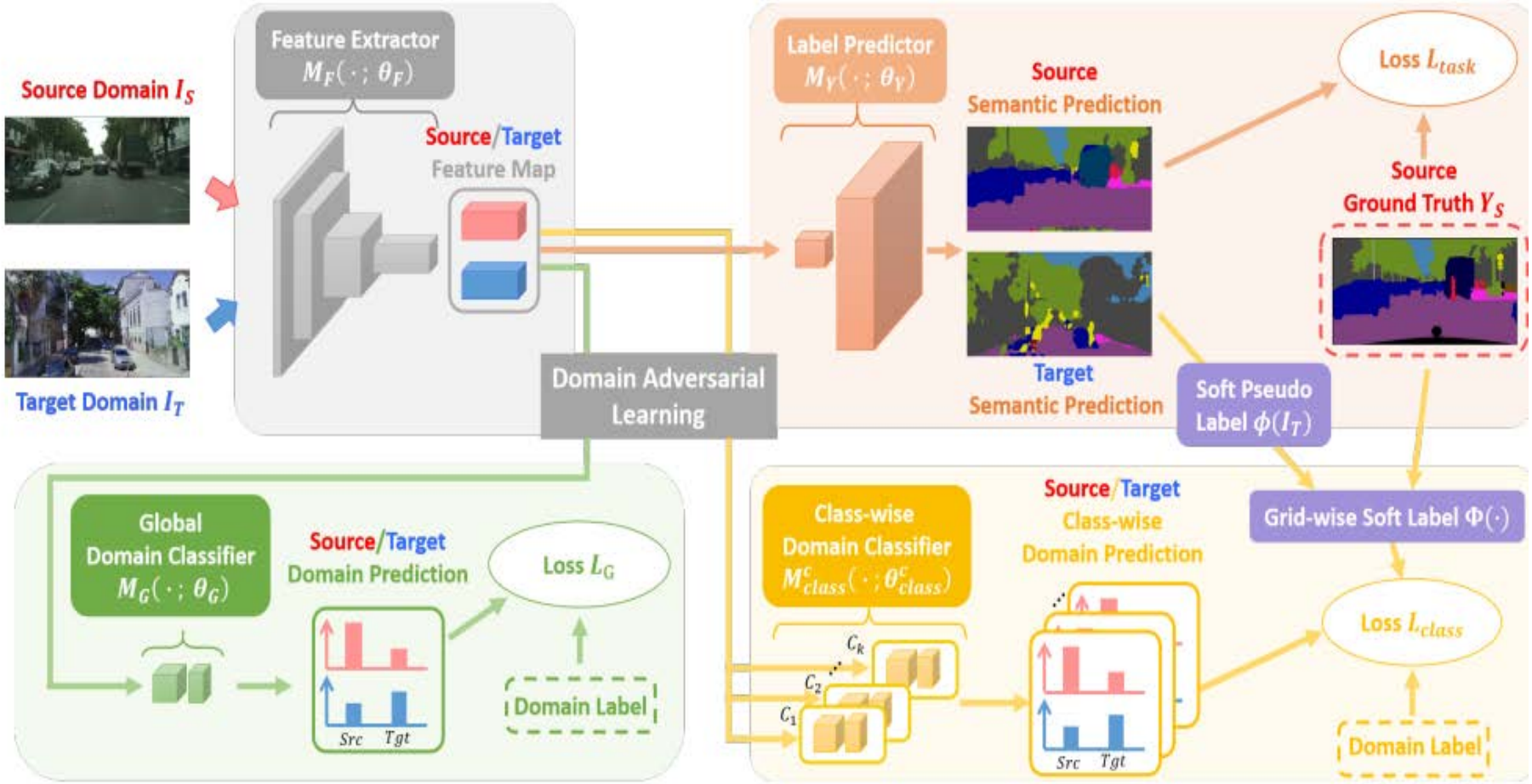
Superpixel



Static classes

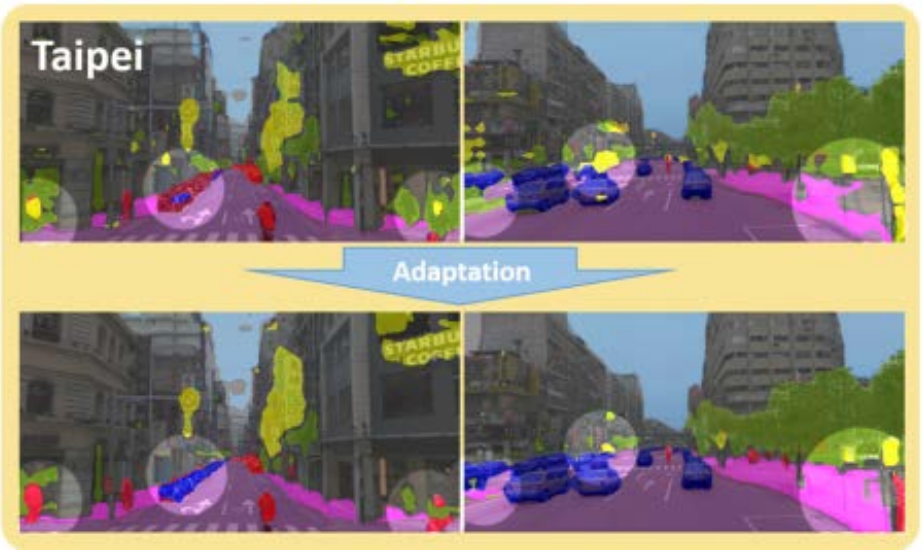
Semantic Segmentation Across Cities

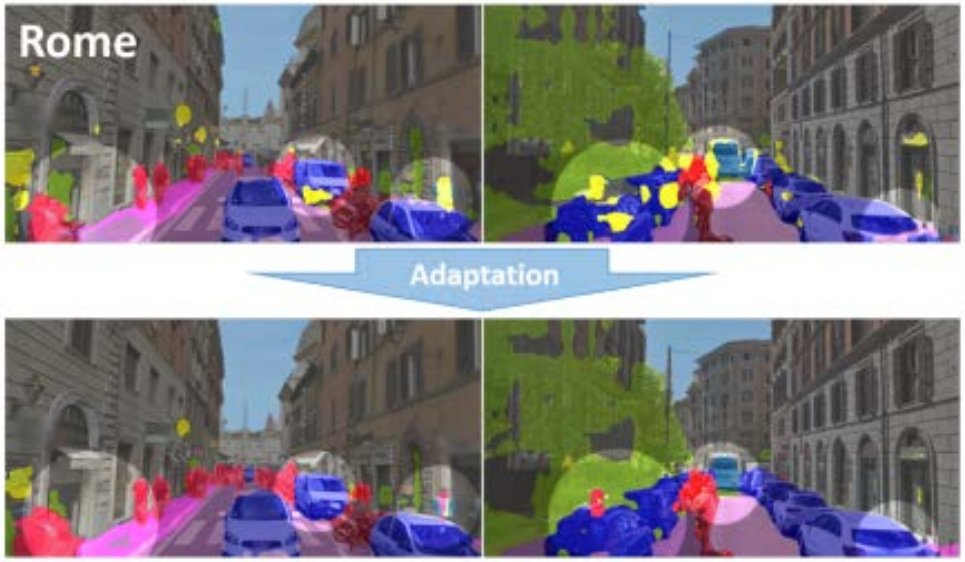
- No More Discrimination: Cross City Adaptation of Road Scene Segmenters



Semantic Segmentation Across Cities

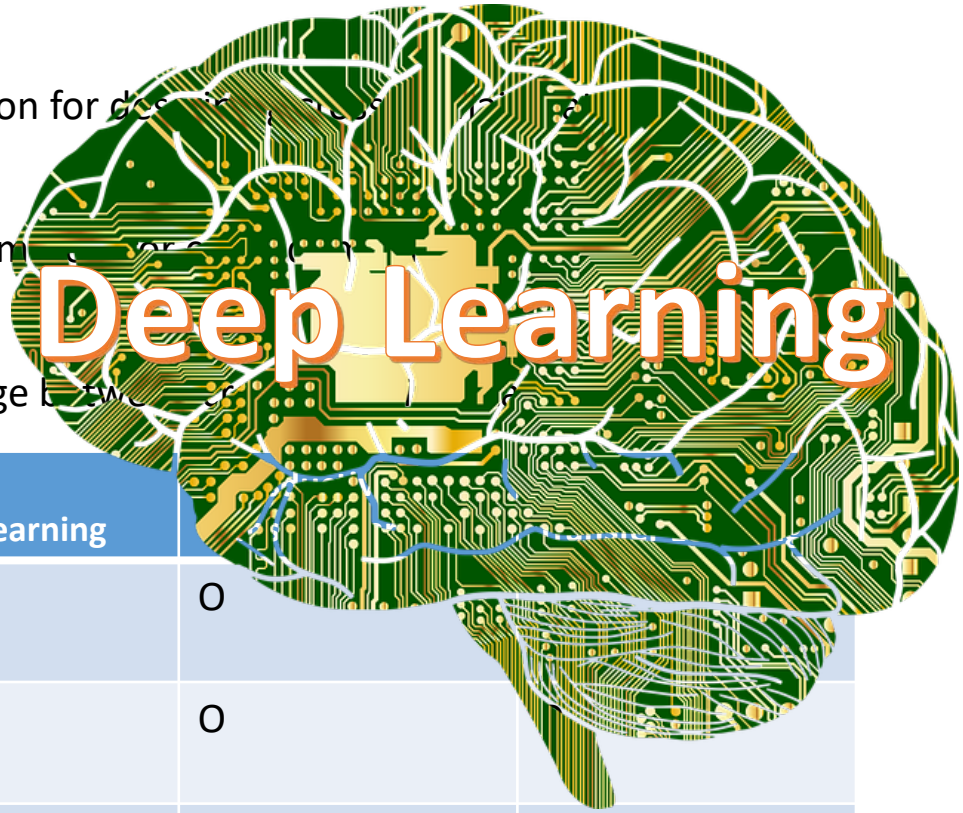
- No More Discrimination: Cross City Adaptation of Road Scene Segmenters
  - Chen et al., ICCV 2017
  - Weakly supervised DA for semantic segmentation
  - Static-object prior from Google Map Time Machine features
  - Qualitative example results





TL Approaches for Cross-Domain Classification

- Instance Transfer
  - Re-weight source-domain label instances for adaptation
- Feature Transfer
  - Derive common feature representation for d[illegible]
- Parameter Transfer
  - Discover shared learning model param[illegible]
- Relational Knowledge Transfer
  - Build mapping of relational knowledge b[illegible]



Methods	Inductive Transfer Learning		
Instance Transfer (Instance Reweighting)	O	O	
Feature Transfer (Common Feature Representation)	O	O	
Parameter/Model Transfer	O		
Relational knowledge Transfer	O		

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

Let's Take a Break…

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