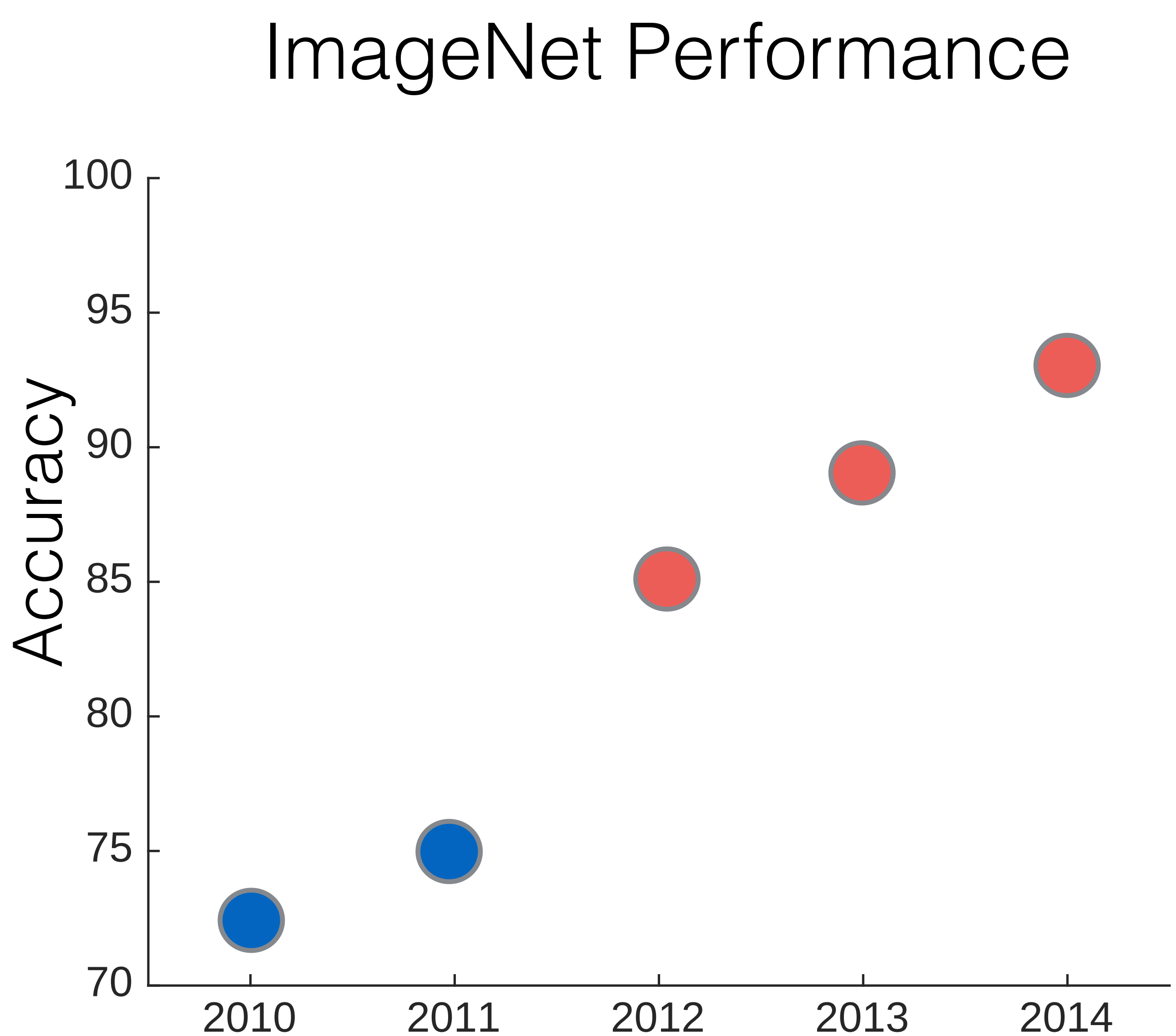
Adapting Deep Networks Across Domains, Modalities, and Tasks

Judy Hoffman

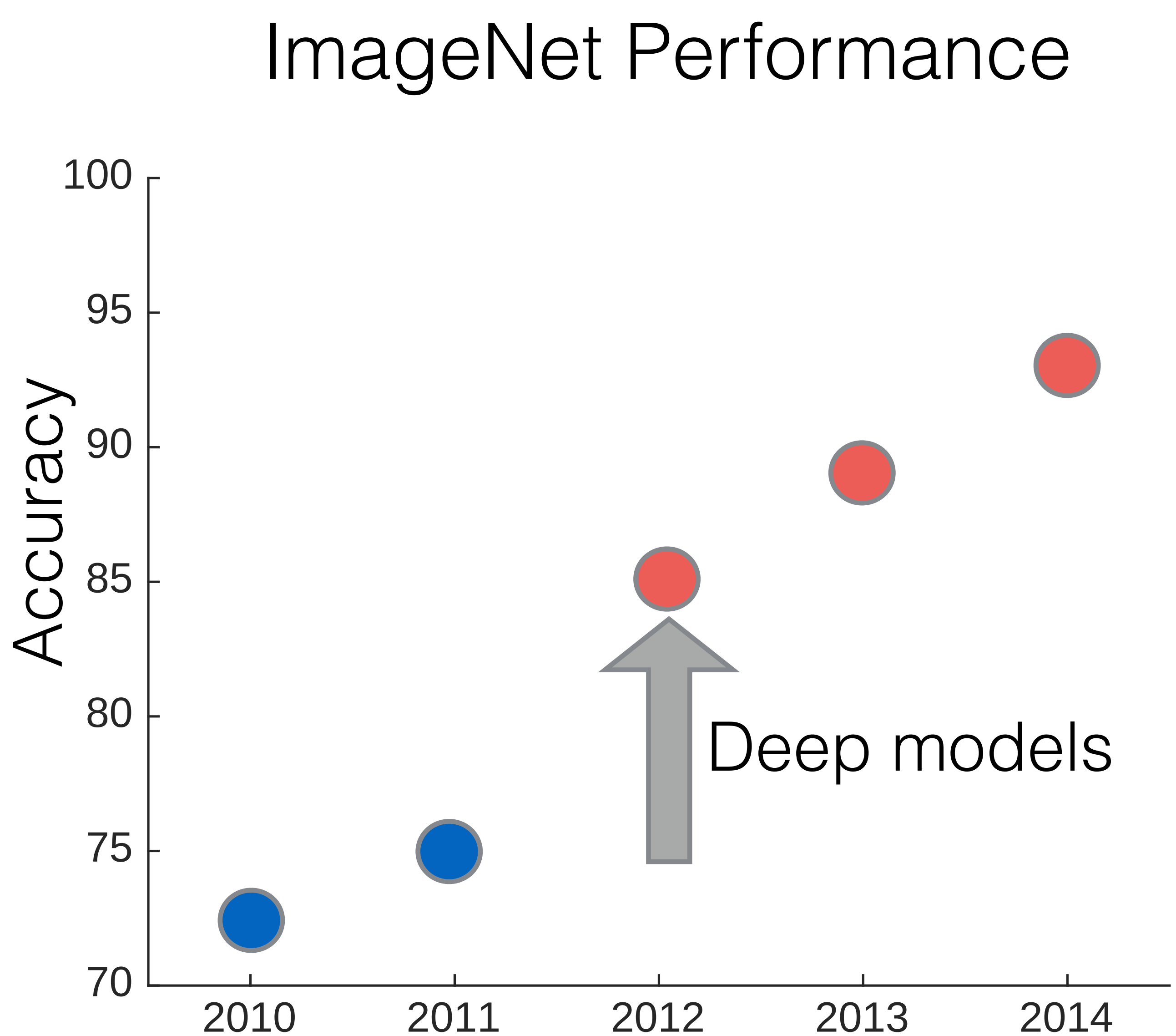



Eric Tzeng | Saurabh Gupta | Kate Saenko | Trevor Darrell

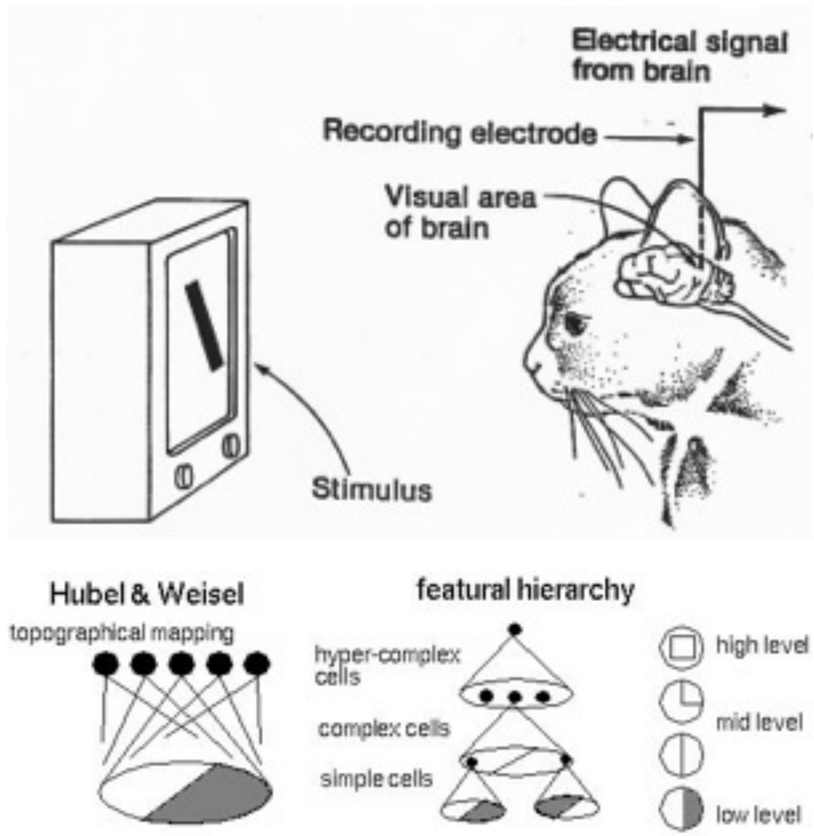
Recent Visual Recognition Progress



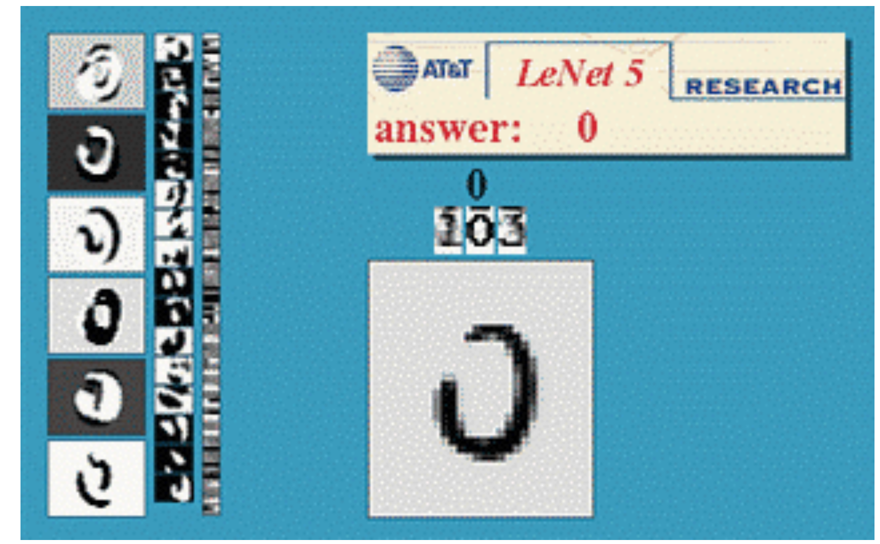
Recent Visual Recognition Progress



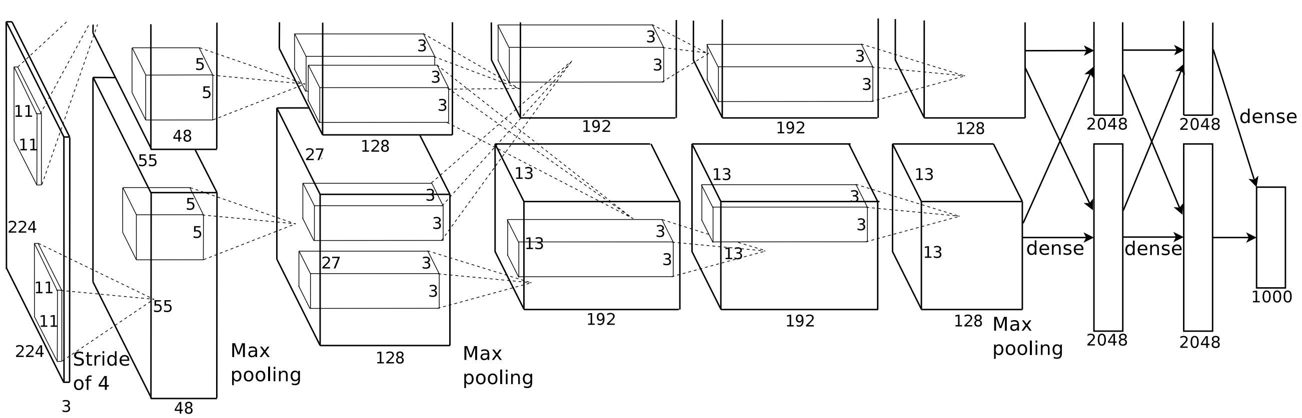
Deep Visual Models



[Hubel and Wisel 59]



[LeCuN 89, 98]

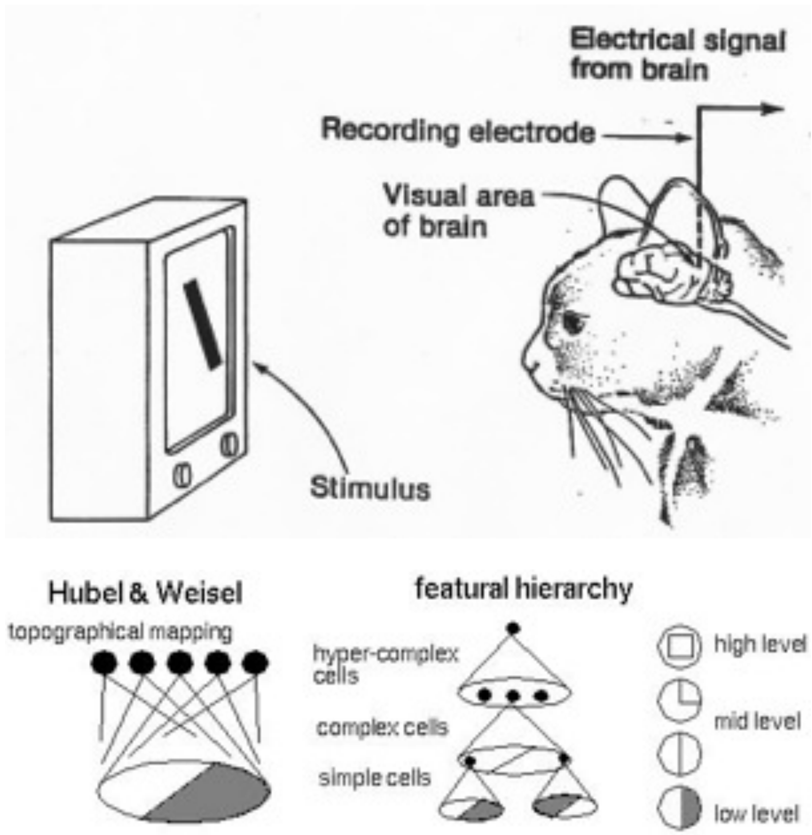


[Krizhevsky 2012]

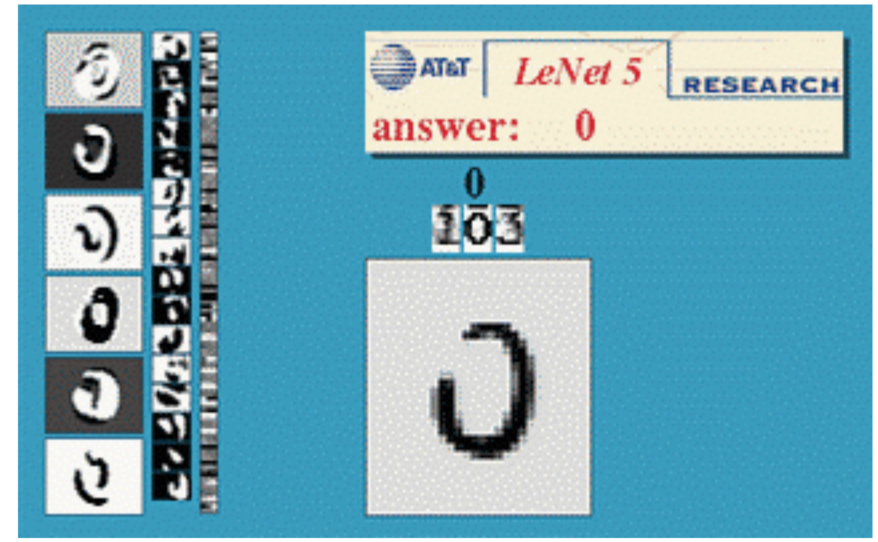
[Szegedy 2014]

[Simonyan 2014]

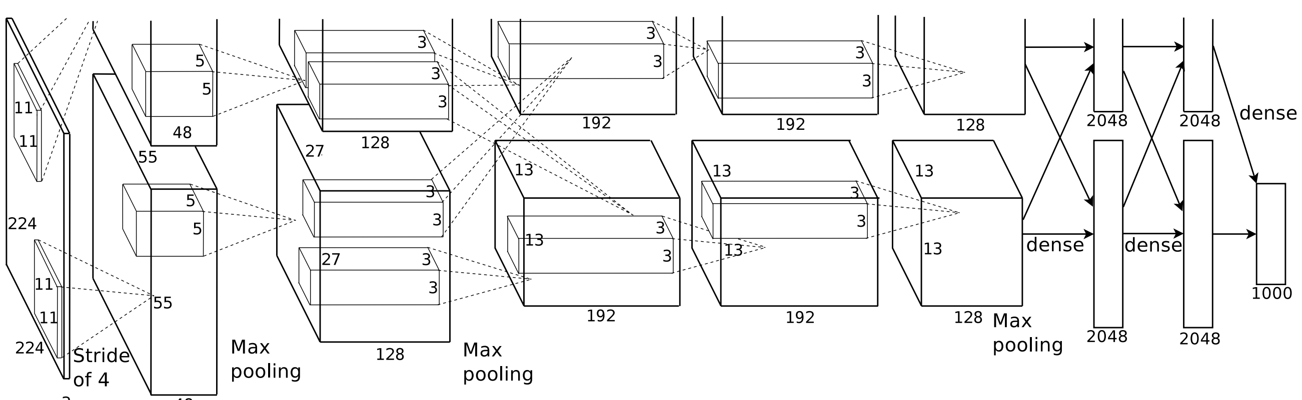
Deep Visual Models



[Hubel and Wisel 59]



[LeCuN 89, 98]



[Krizhevsky 2012]

[Szegedy 2014]

[Simonyan 2014]

Domain Adaptation: Train on source adapt to target

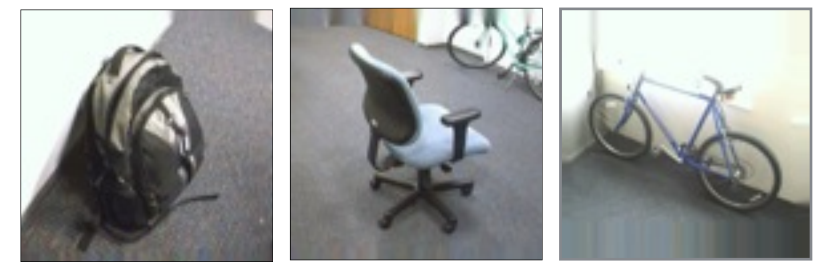
Source Domain $\sim P_S(X, Y)$

lots of **labeled** data

$$D_S = \{(\mathbf{x}_i, y_i), \forall i \in \{1, \dots, N\}\}$$

Domain Adaptation: Train on source adapt to target

backpack · · · chair bike



? ? bike

Source Domain $\sim P_S(X, Y)$ $\neq$ Target Domain $\sim P_T(Z, H)$

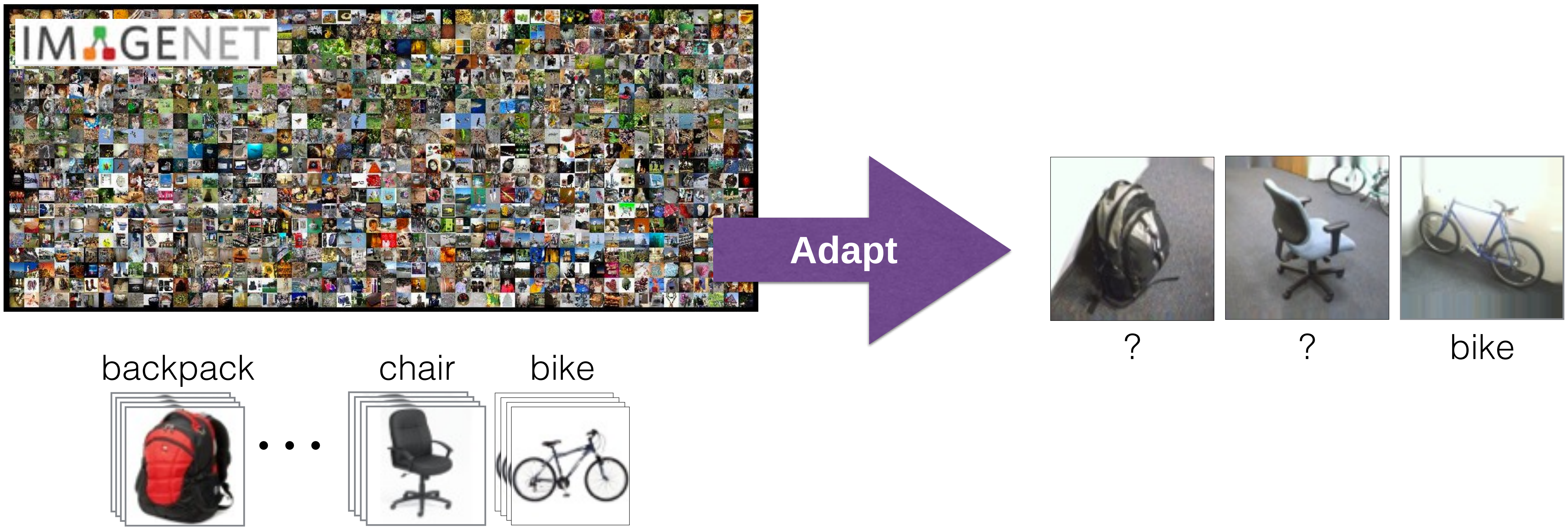
lots of **labeled** data

unlabeled or limited labels

$$D_S = \{(\mathbf{x}_i, y_i), \forall i \in \{1, \dots, N\}\}$$

$$D_T = \{(\mathbf{z}_j, \mathbf{?}), \forall j \in \{1, \dots, M\}\}$$

Domain Adaptation: Train on source adapt to target



Source Domain $\sim P_S(X, Y)$ $\neq$ Target Domain $\sim P_T(Z, H)$

lots of **labeled** data

unlabeled or limited labels

$$D_S = \{(\mathbf{x}_i, y_i), \forall i \in \{1, \dots, N\}\}$$

$$D_T = \{(\mathbf{z}_j, \mathbf{?}), \forall j \in \{1, \dots, M\}\}$$

Prior work: domain adaptation

- Minimizing distribution distance
  - Borgwardt`06, Mansour`09, Pan`09, Fernando`13
- Deep model adaptation
  - Chopra`13, Tzeng`14, Long`15, Ganin`15

Adapting across domains minimize discrepancy

$$\min_{\theta_{\mathrm{repr}}} \mathcal{L}_{\mathrm{conf}}(x, z, \theta_D; \theta_{\mathrm{repr}}) = \sum_{x_i \in S} H(\mathcal{U}(D), q_i^s) + \sum_{z_j \in T} H(\mathcal{U}(D), q_j^t)$$

[ICCV 2015]

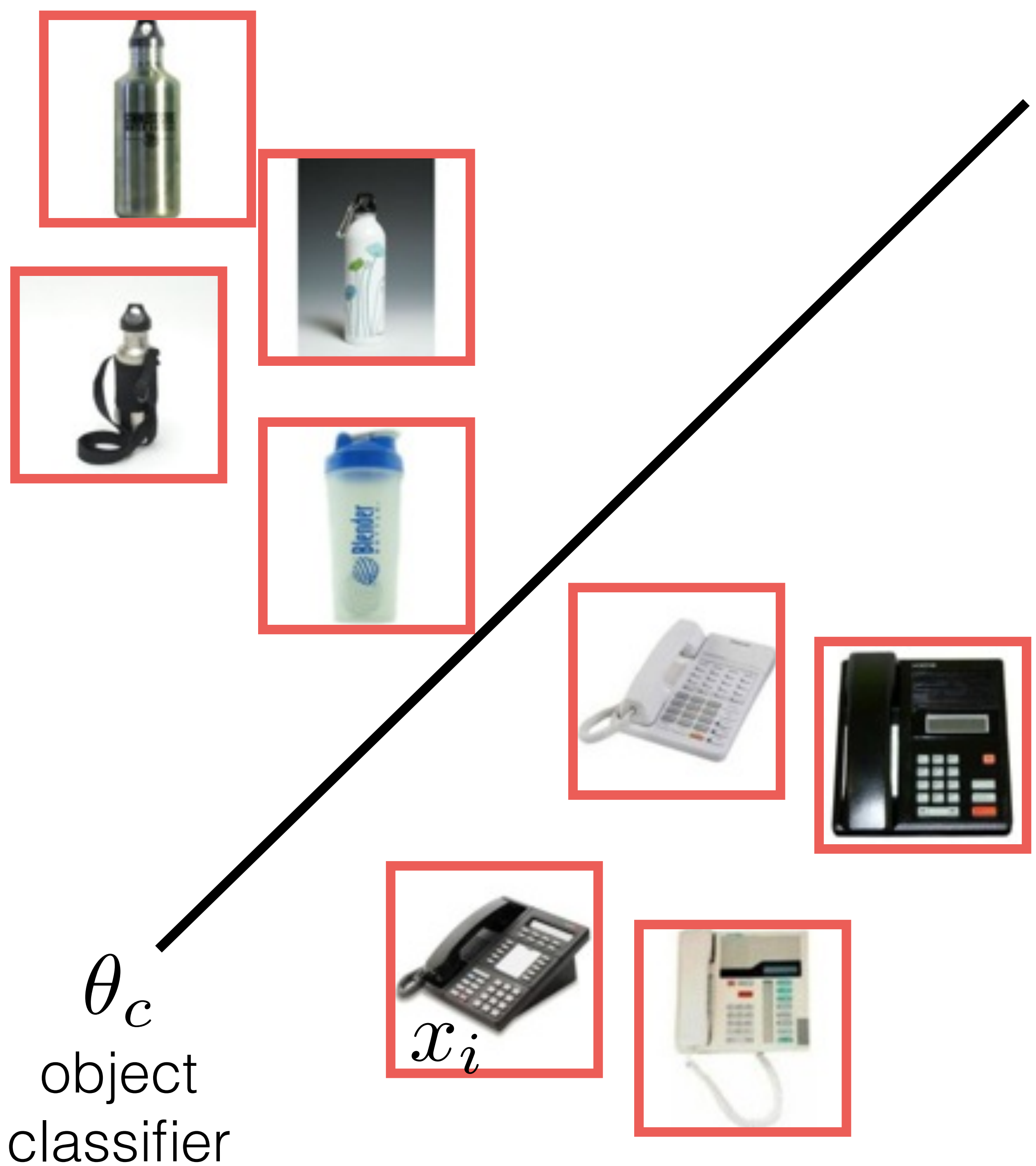
Adapting across domains minimize discrepancy

$$\min_{\theta_{\mathrm{repr}}} \mathcal{L}_{\mathrm{conf}}(x, z, \theta_D; \theta_{\mathrm{repr}}) = \sum_{x_i \in S} H(\mathcal{U}(D), q_i^s) + \sum_{z_j \in T} H(\mathcal{U}(D), q_j^t)$$

[ICCV 2015]

Adapting across domains minimize discrepancy

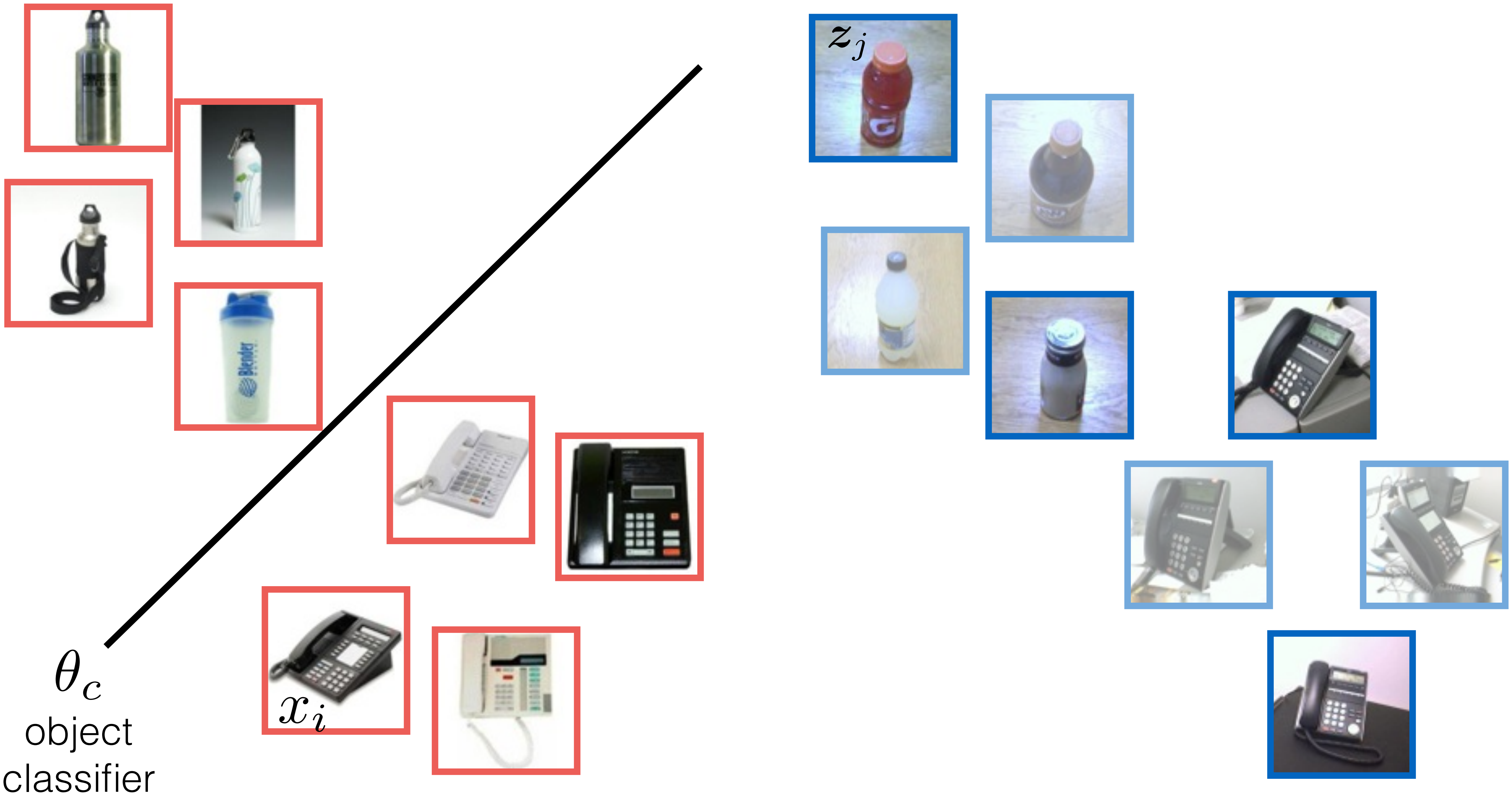
$$\min_{\theta_{\text{repr}}} \mathcal{L}_{\text{conf}}(x, z, \theta_D; \theta_{\text{repr}}) = \sum_{x_i \in S} H(\mathcal{U}(D), q_i^s) + \sum_{z_j \in T} H(\mathcal{U}(D), q_j^t)$$



[ICCV 2015]

Adapting across domains minimize discrepancy

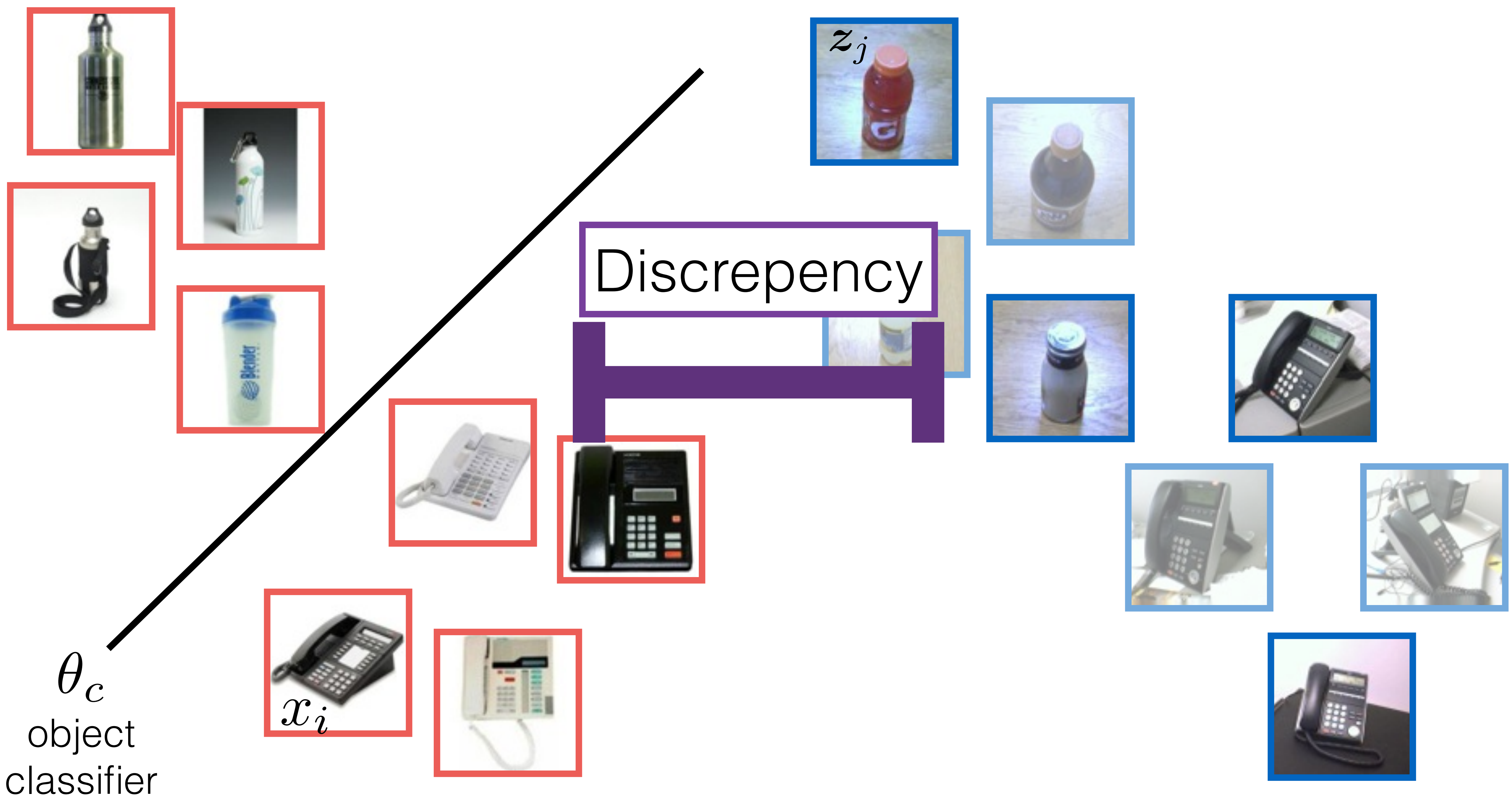
$$\min_{\theta_{\text{repr}}} \mathcal{L}_{\text{conf}}(x, z, \theta_D; \theta_{\text{repr}}) = \sum_{x_i \in S} H(\mathcal{U}(D), q_i^s) + \sum_{z_j \in T} H(\mathcal{U}(D), q_j^t)$$



[ICCV 2015]

Adapting across domains minimize discrepancy

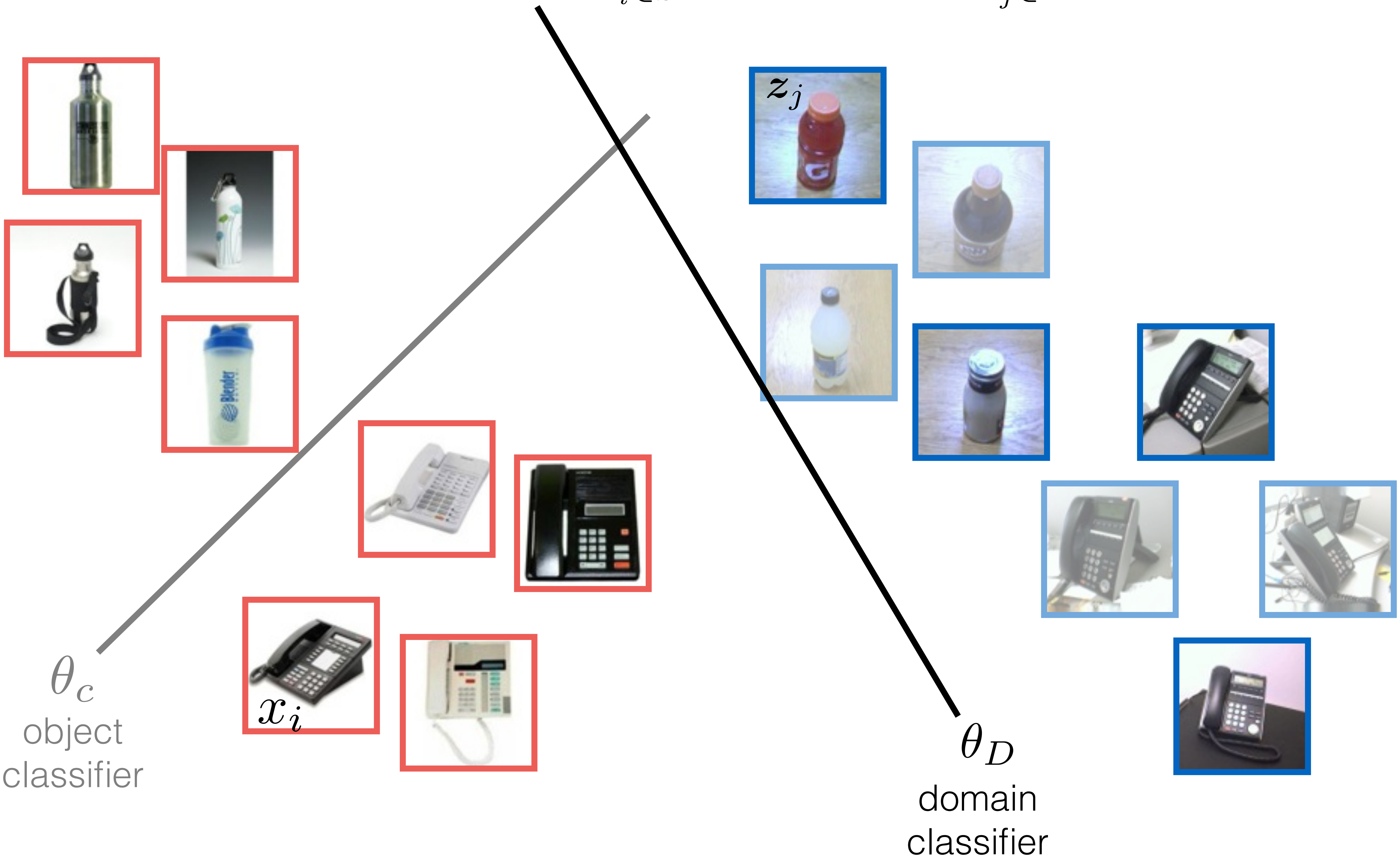
$$\min_{\theta_{\text{repr}}} \mathcal{L}_{\text{conf}}(x, z, \theta_D; \theta_{\text{repr}}) = \sum_{x_i \in S} H(\mathcal{U}(D), q_i^s) + \sum_{z_j \in T} H(\mathcal{U}(D), q_j^t)$$



[ICCV 2015]

Adapting across domains minimize discrepancy

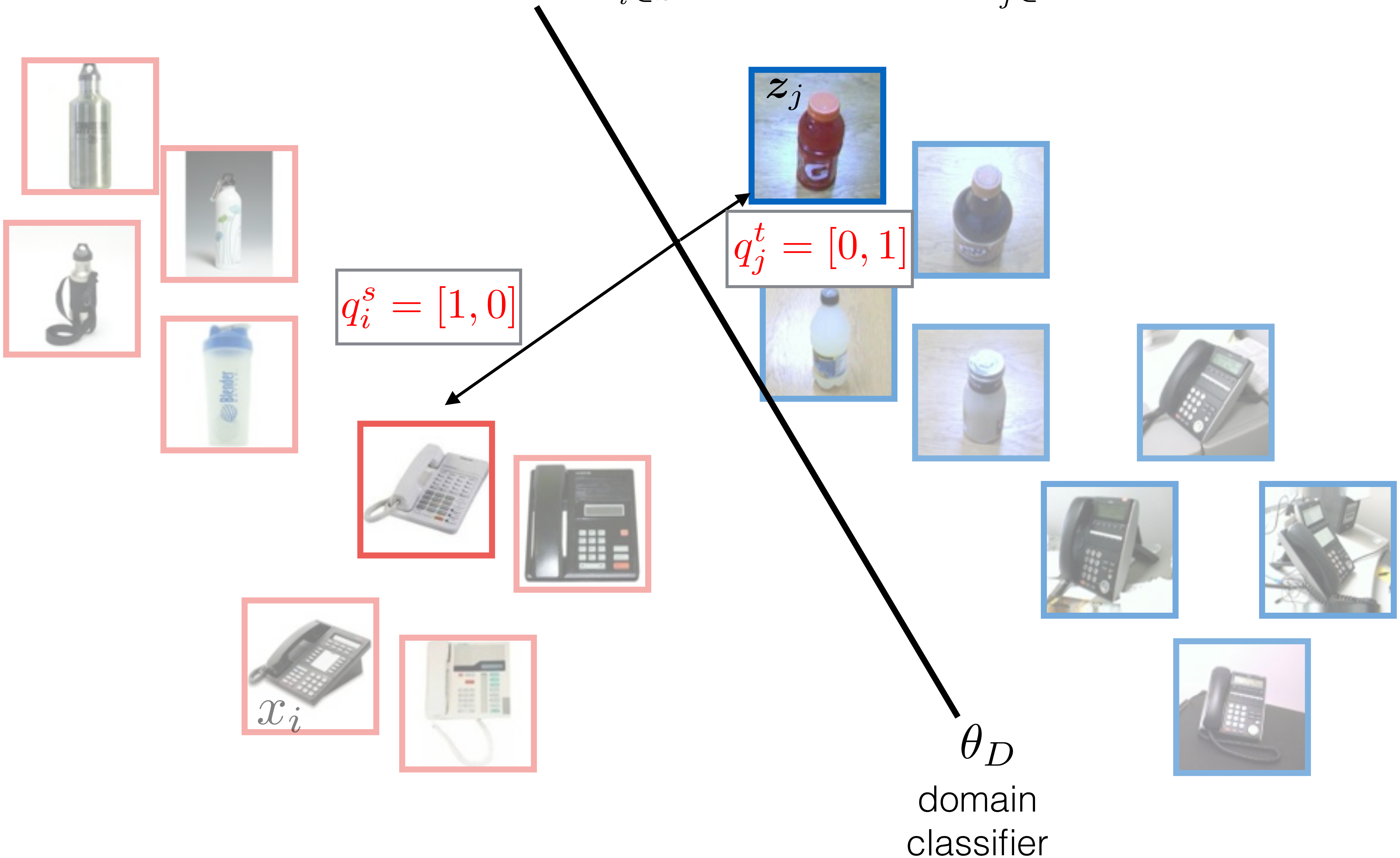
$$\min_{\theta_{\text{repr}}} \mathcal{L}_{\text{conf}}(x, z, \theta_D; \theta_{\text{repr}}) = \sum_{x_i \in S} H(\mathcal{U}(D), q_i^s) + \sum_{z_j \in T} H(\mathcal{U}(D), q_j^t)$$



[ICCV 2015]

Adapting across domains minimize discrepancy

$$\min_{\theta_{\text{repr}}} \mathcal{L}_{\text{conf}}(x, z, \theta_D; \theta_{\text{repr}}) = \sum_{x_i \in S} H(\mathcal{U}(D), q_i^s) + \sum_{z_j \in T} H(\mathcal{U}(D), q_j^t)$$

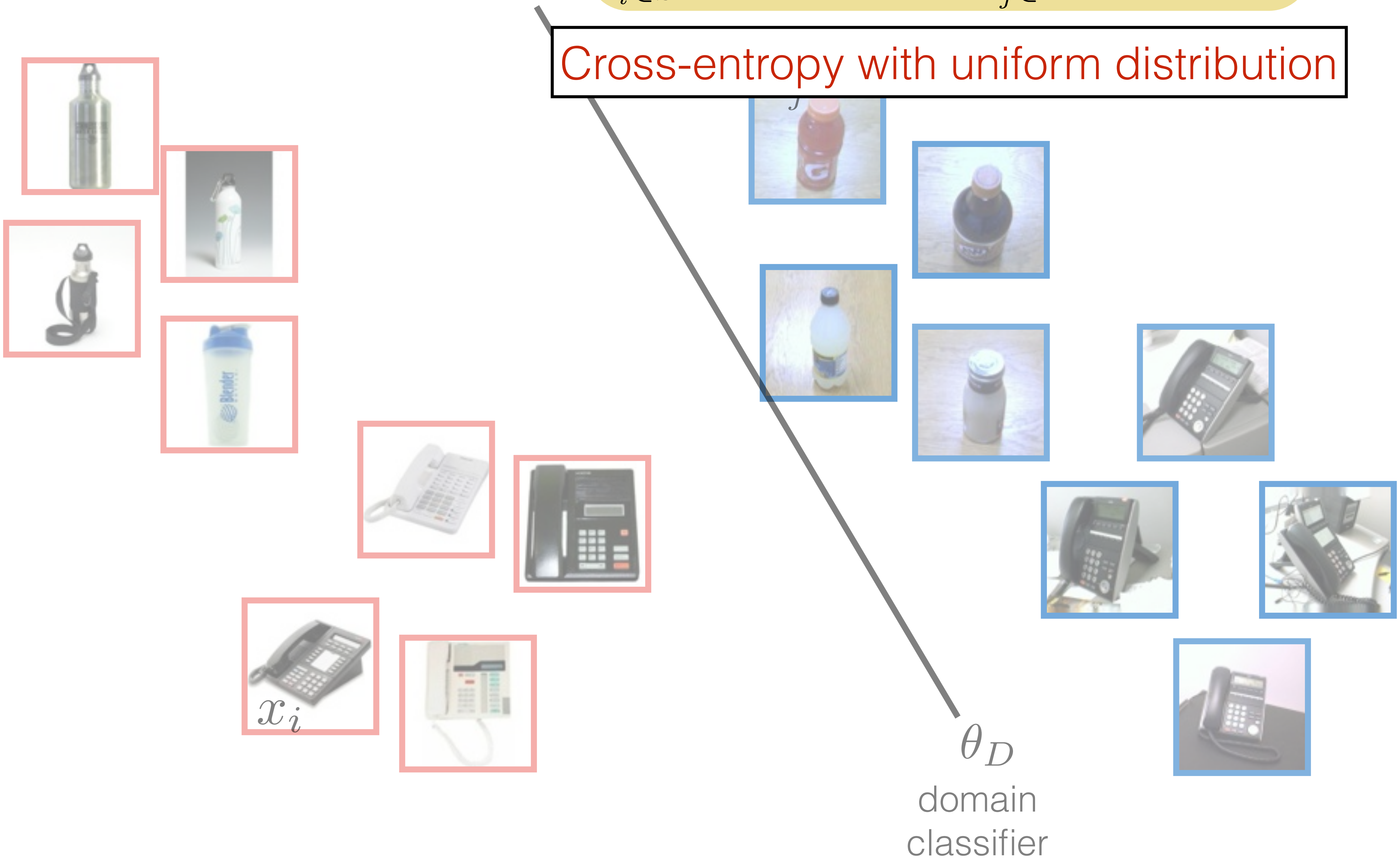


[ICCV 2015]

Adapting across domains minimize discrepancy

$$\min_{\theta_{\text{repr}}} \mathcal{L}_{\text{conf}}(x, z, \theta_D; \theta_{\text{repr}}) = \sum_{x_i \in S} H(\mathcal{U}(D), q_i^s) + \sum_{z_j \in T} H(\mathcal{U}(D), q_j^t)$$

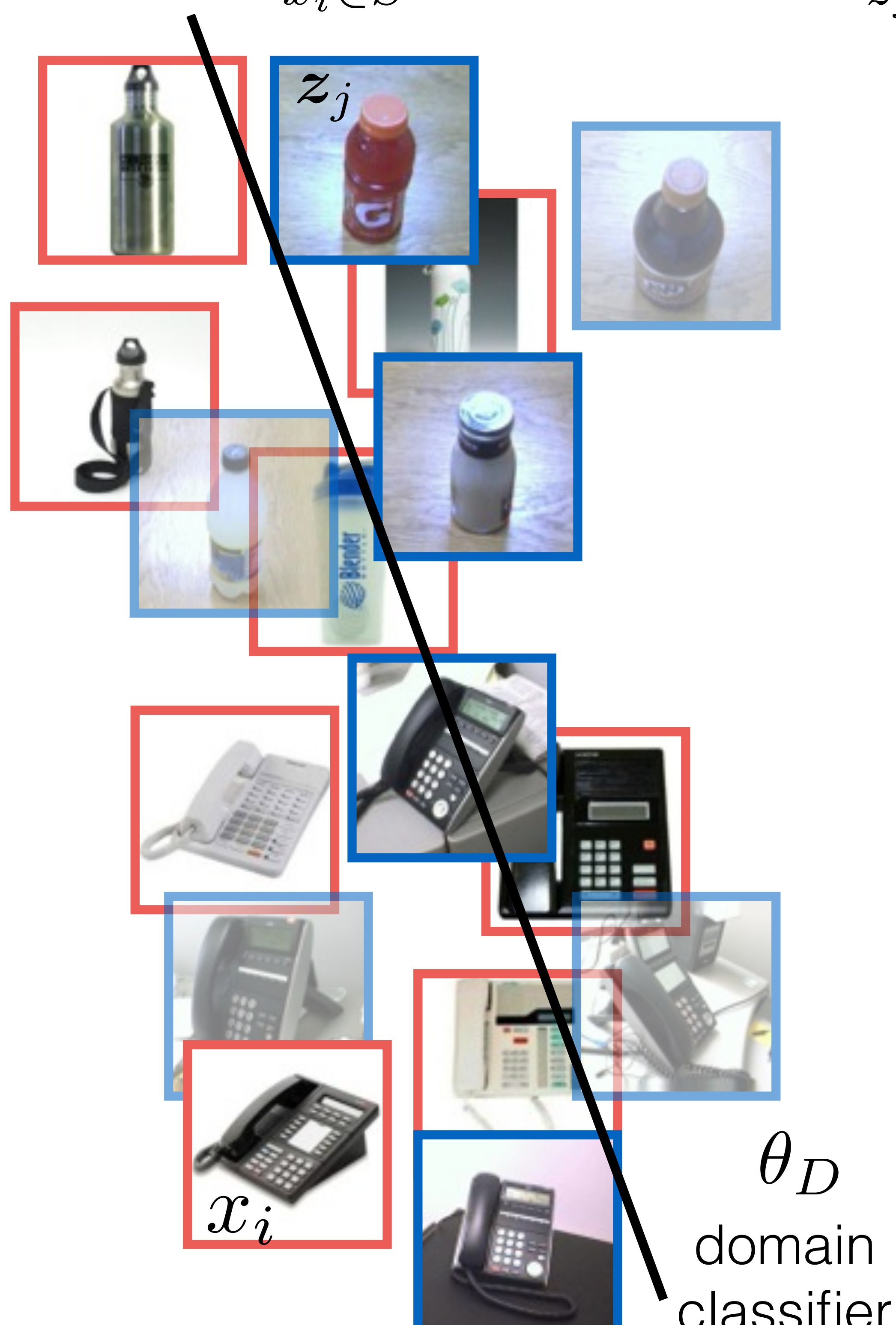
Cross-entropy with uniform distribution



[ICCV 2015]

Adapting across domains minimize discrepancy

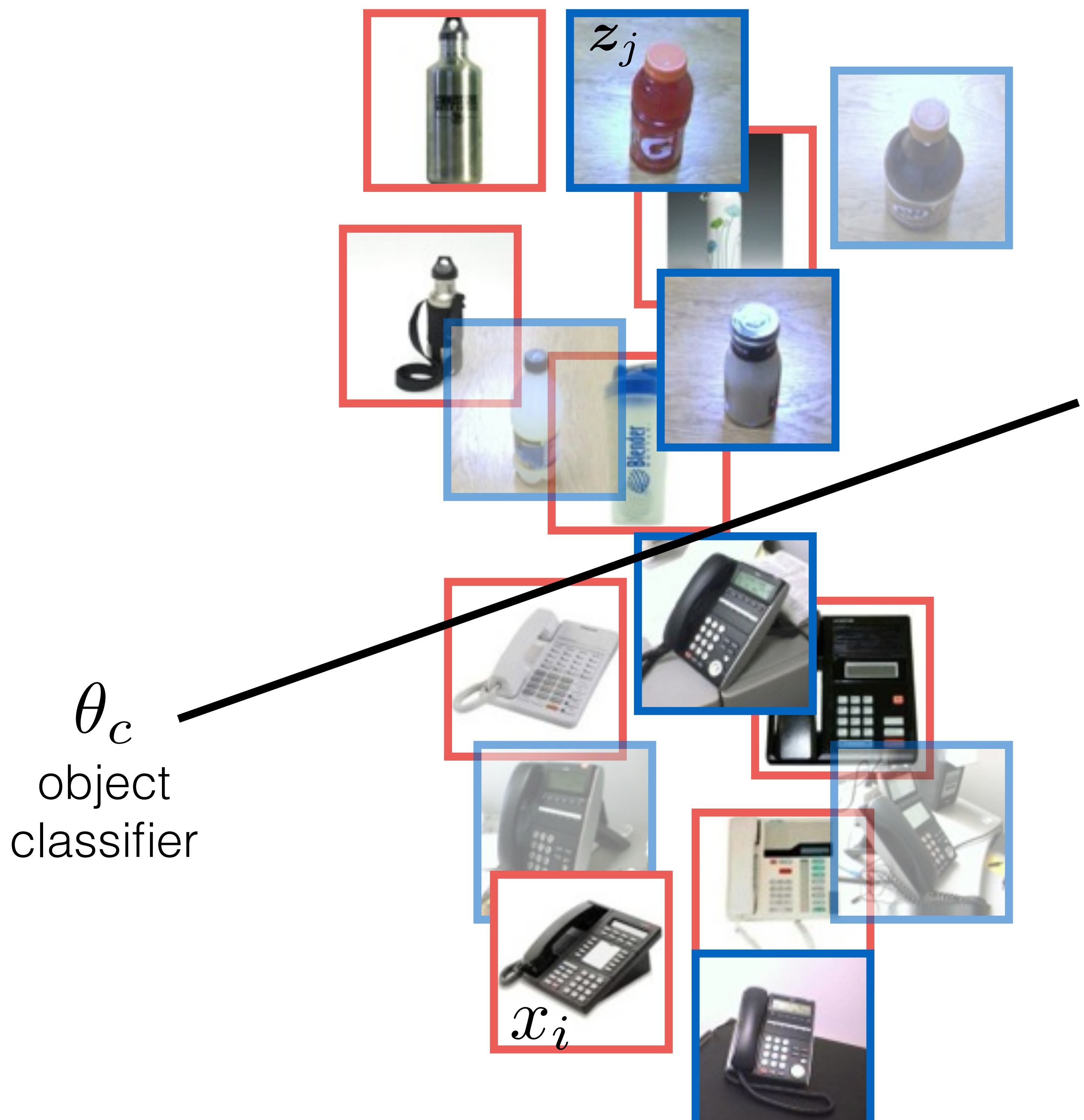
$$\min_{\theta_{\text{repr}}} \mathcal{L}_{\text{conf}}(x, z, \theta_D; \theta_{\text{repr}}) = \sum_{x_i \in S} H(\mathcal{U}(D), q_i^s) + \sum_{z_j \in T} H(\mathcal{U}(D), q_j^t)$$



[ICCV 2015]

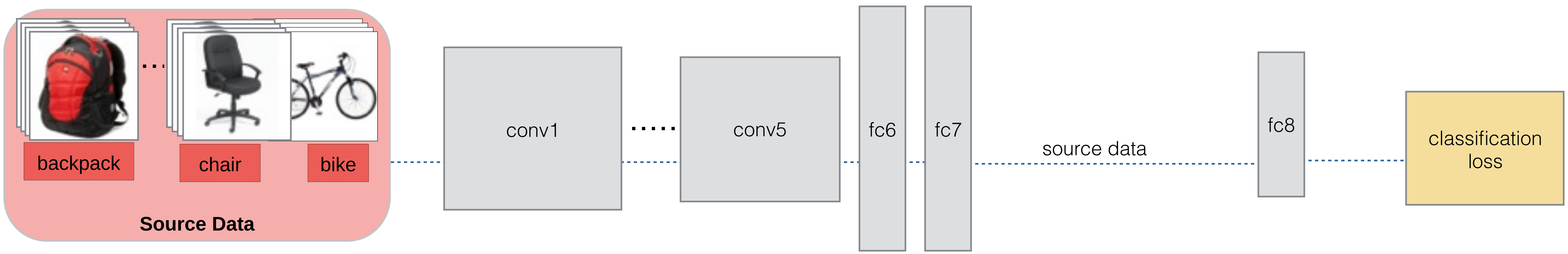
Adapting across domains minimize discrepancy

$$\min_{\theta_{\text{repr}}} \mathcal{L}_{\text{conf}}(x, z, \theta_D; \theta_{\text{repr}}) = \sum_{x_i \in S} H(\mathcal{U}(D), q_i^s) + \sum_{z_j \in T} H(\mathcal{U}(D), q_j^t)$$

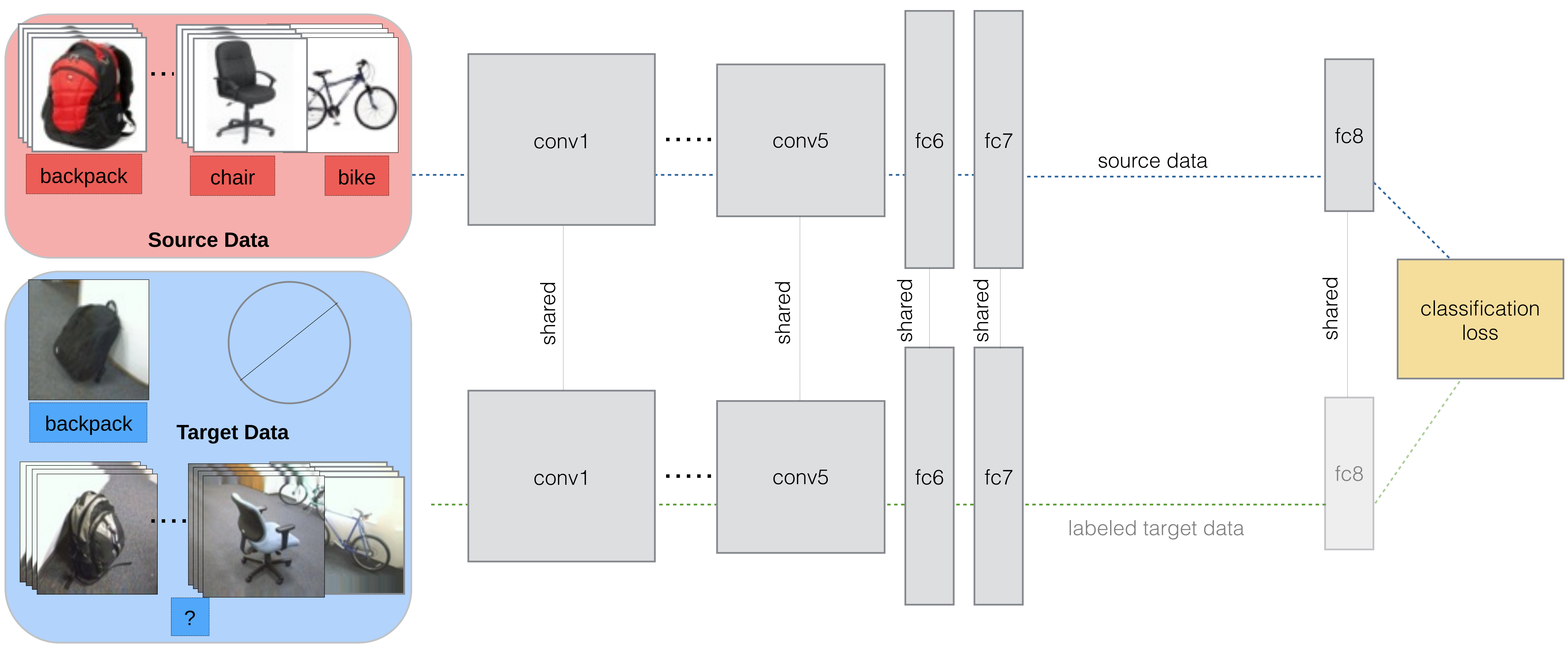


[ICCV 2015]

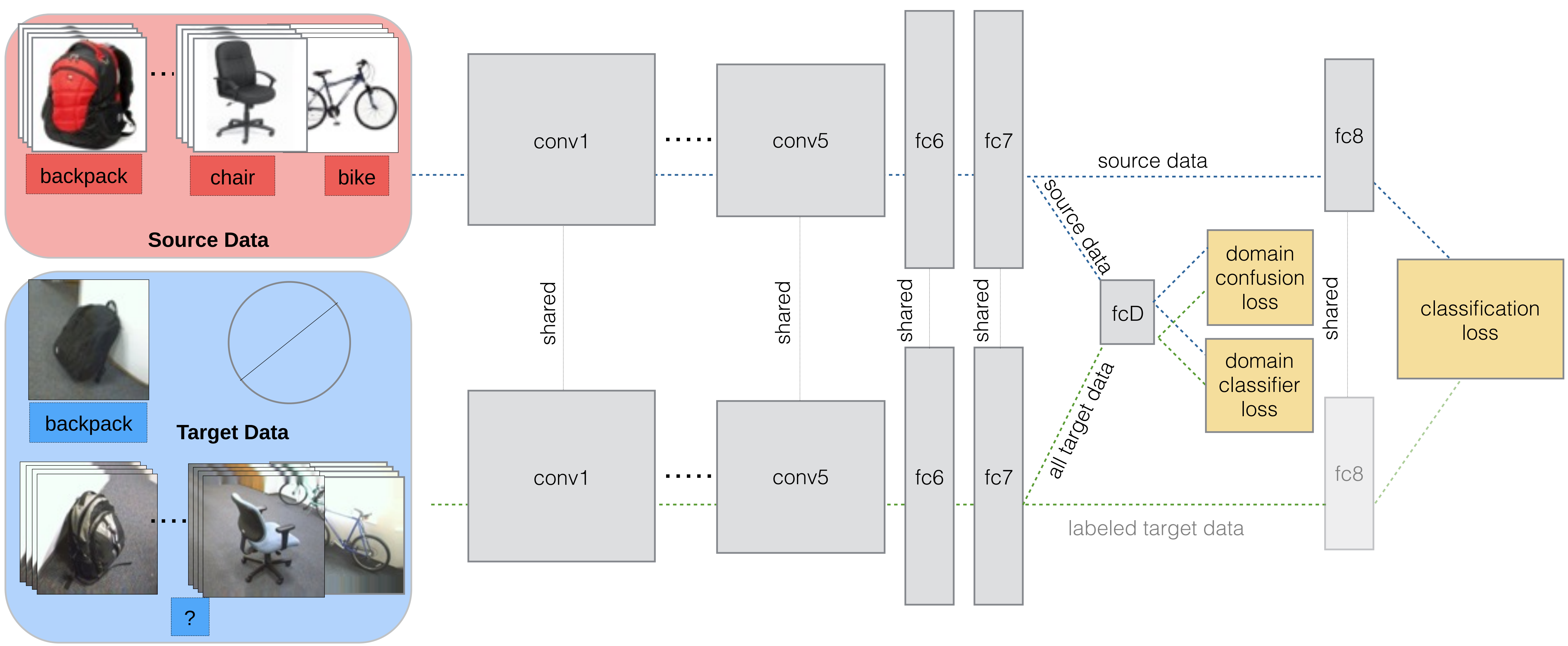
Adapting across domains minimize discrepancy



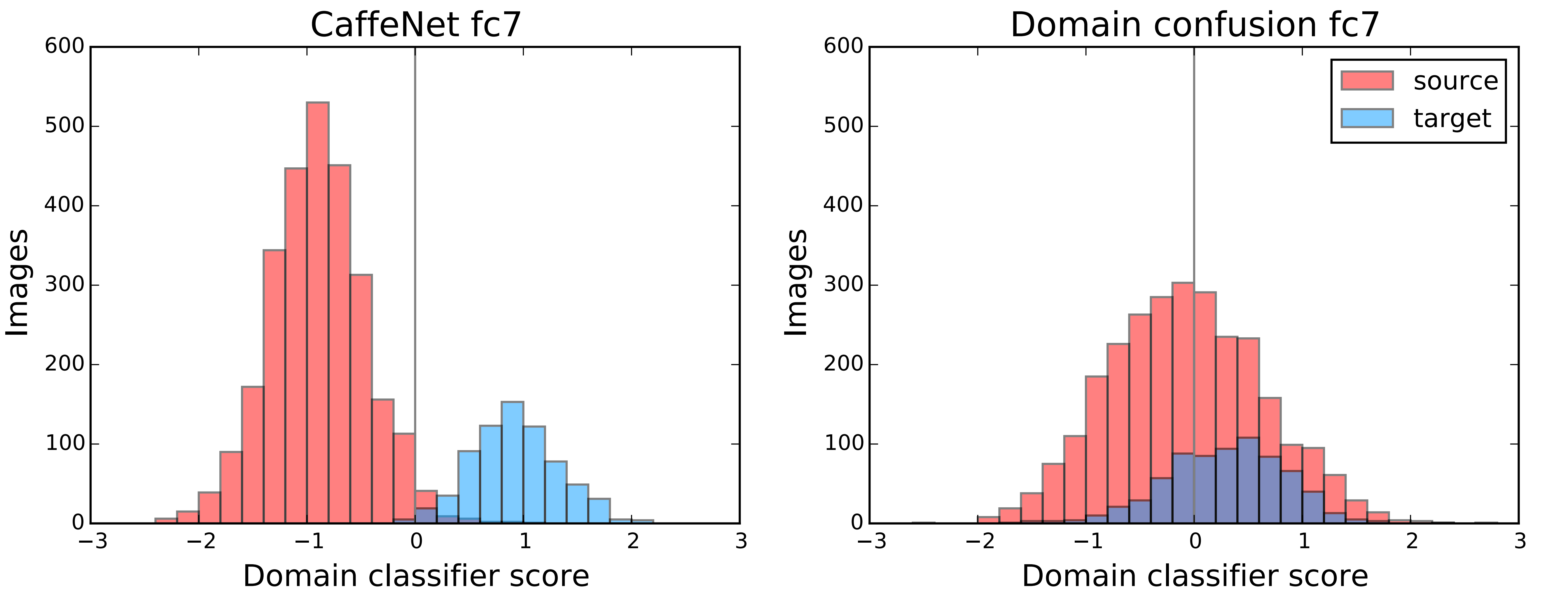
Adapting across domains minimize discrepancy



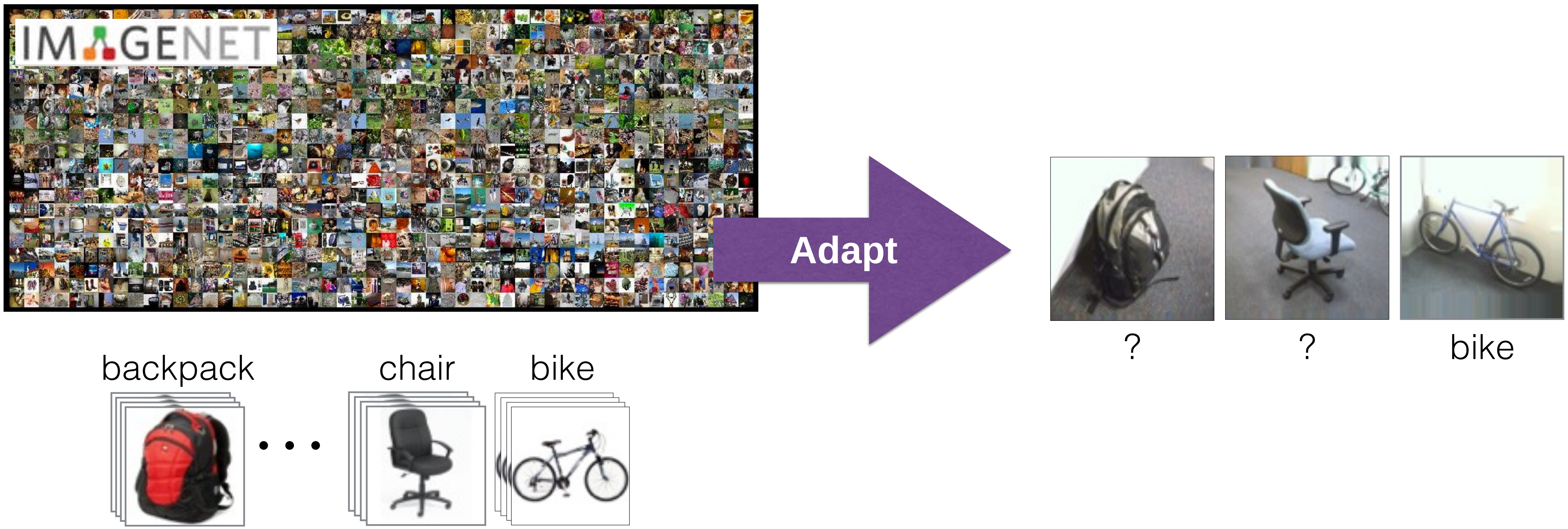
Adapting across domains minimize discrepancy



Verify confusion



Domain Adaptation: Train on source adapt to target



Source Domain $\sim P_S(X, Y)$ $\neq$ Target Domain $\sim P_T(Z, H)$

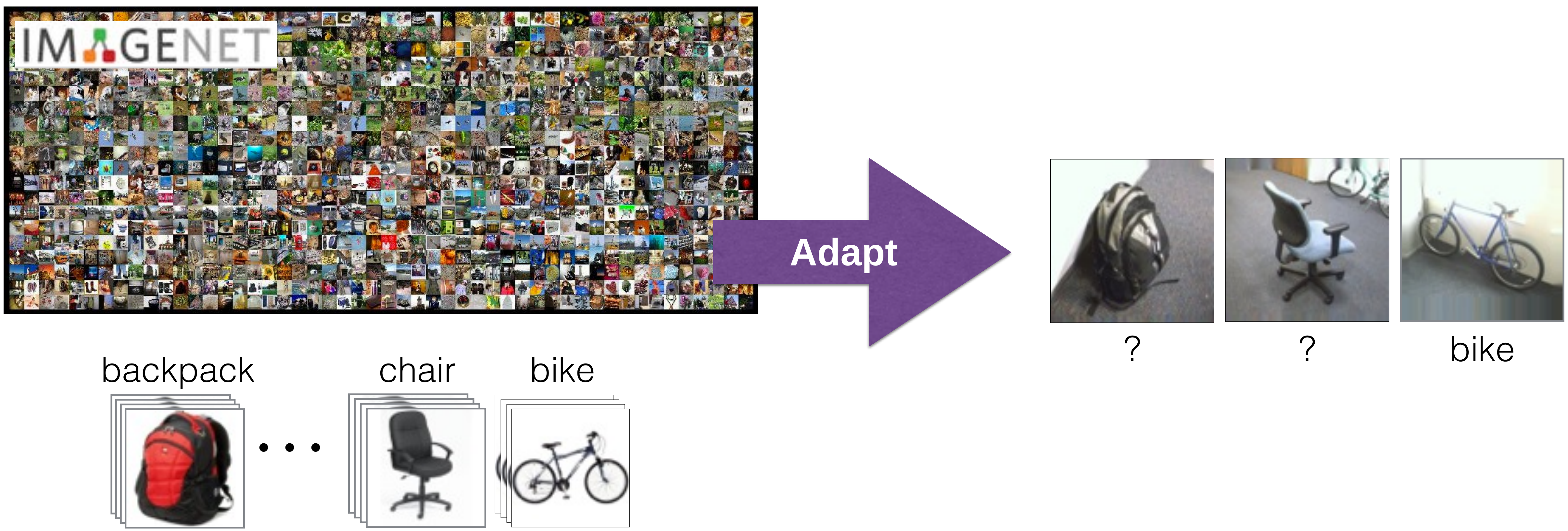
lots of **labeled** data

unlabeled or limited labels

$$D_S = \{(\mathbf{x}_i, y_i), \forall i \in \{1, \ldots, N\}\}$$

$$D_T = \{(\mathbf{z}_j, \mathbf{?}), \forall j \in \{1, \ldots, M\}\}$$

Domain Adaptation: Train on source adapt to target



Source Domain $\sim P_S(X, Y)$

lots of **labeled** data

$$\neq$$

Target Domain $\sim P_T(Z, H)$

unlabeled or limited labels

$$D_S = \{(\mathbf{x}_i, y_i), \forall i \in \{1, \ldots, N\}\}$$

$$D_T = \{(\mathbf{z}_j, \mathbf{?}), \forall j \in \{1, \ldots, M\}\}$$

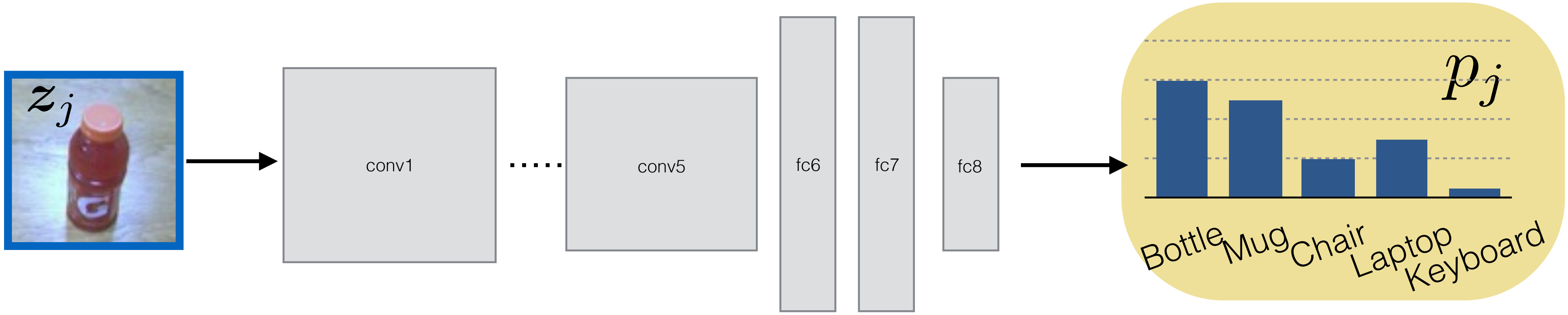
Standard supervised deep loss

$$H(h, p) = E_h[-\log p]$$

[ICCV 2015]

Standard supervised deep loss

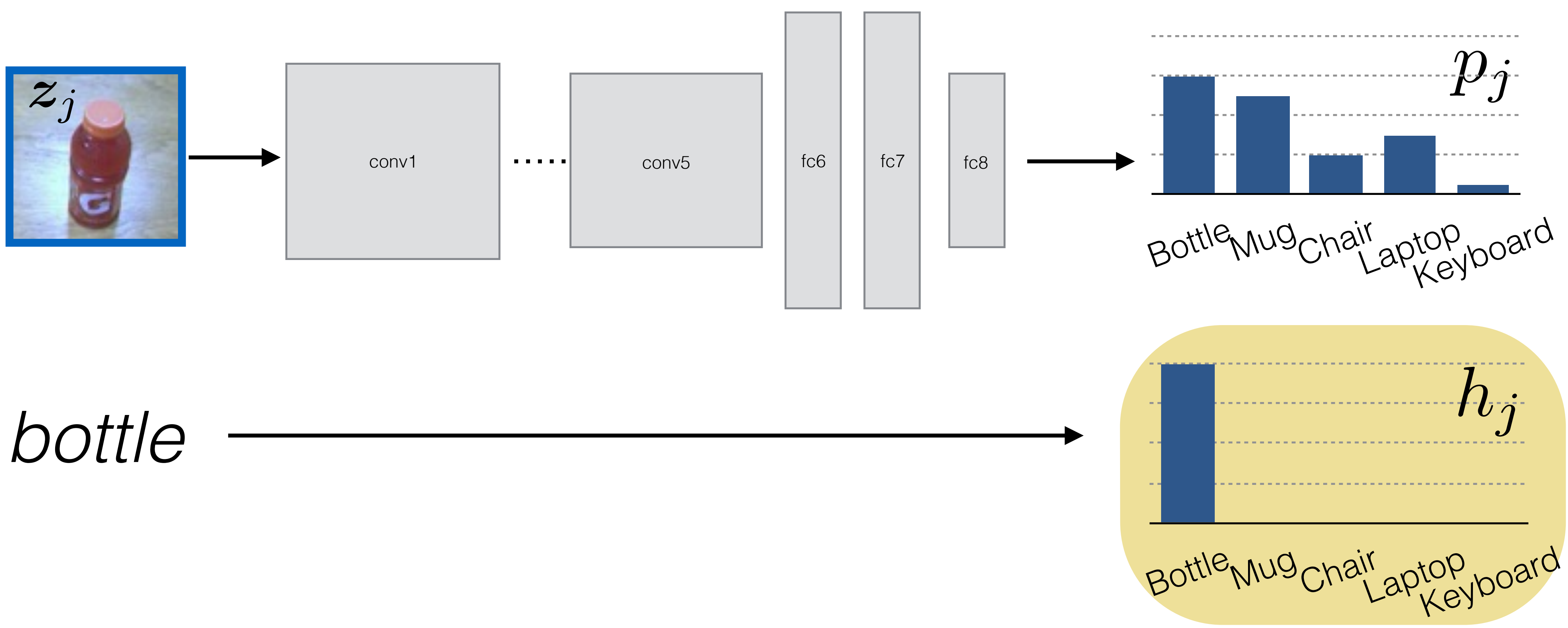
$$H(h, p) = E_h[-\log p]$$



[ICCV 2015]

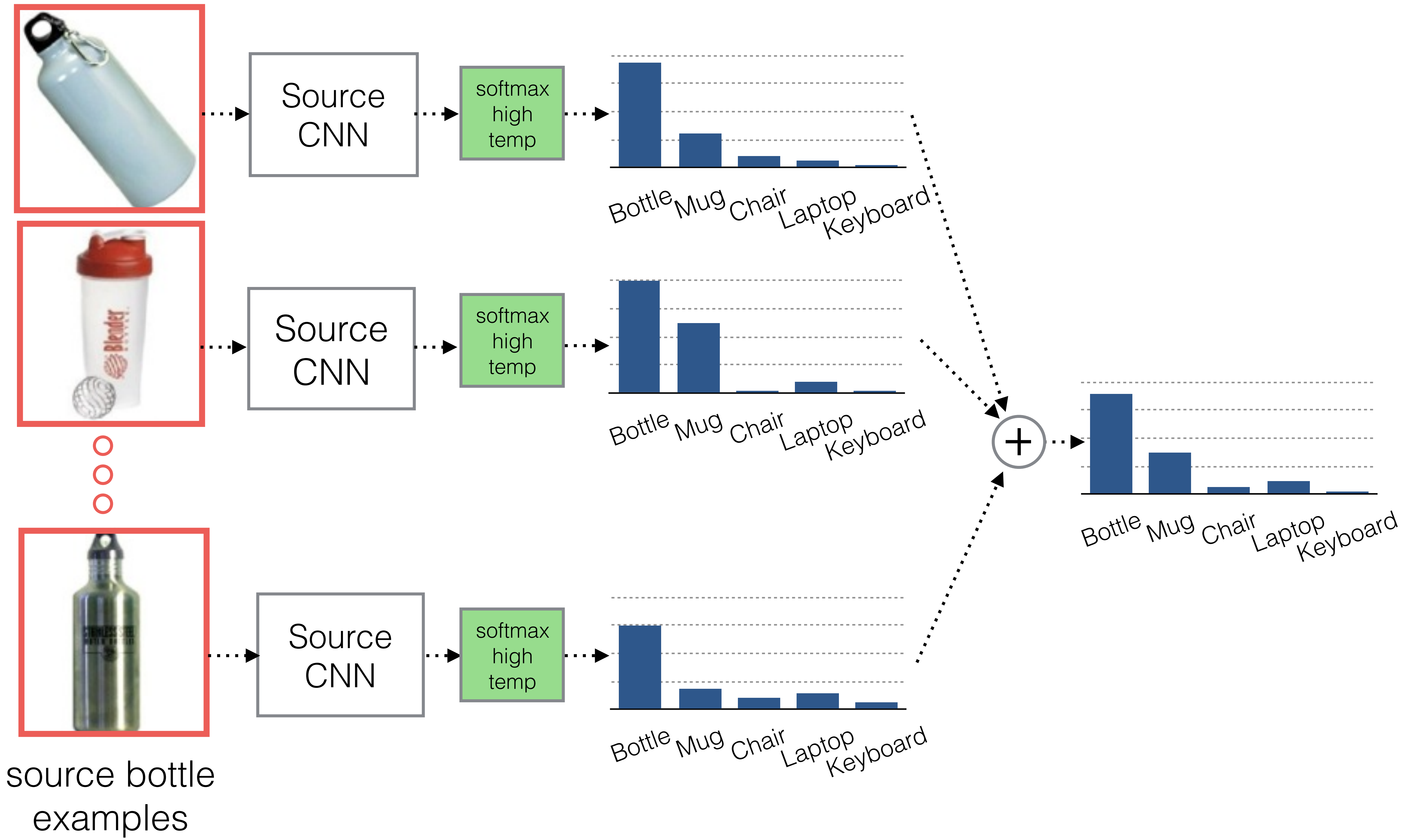
Standard supervised deep loss

$$H(h, p) = E_h[-\log p]$$



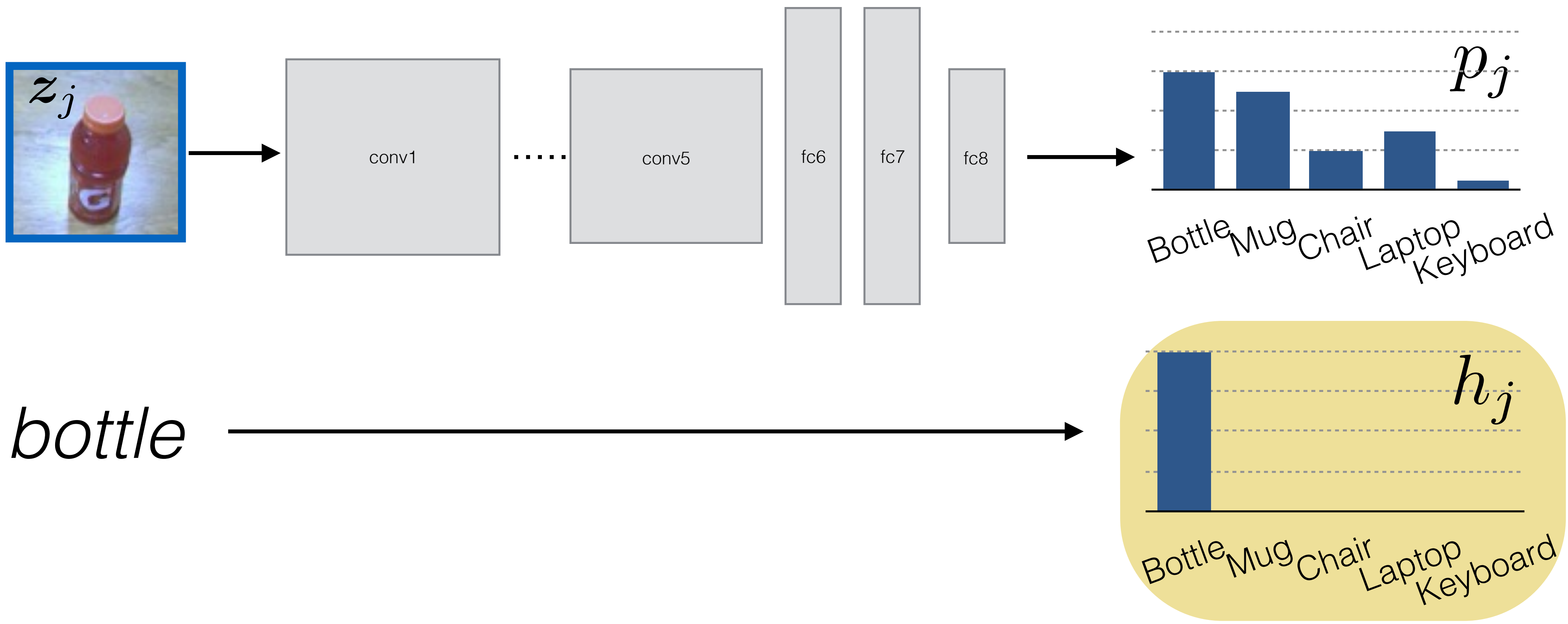
[ICCV 2015]

Source Softlabels



Standard supervised deep loss

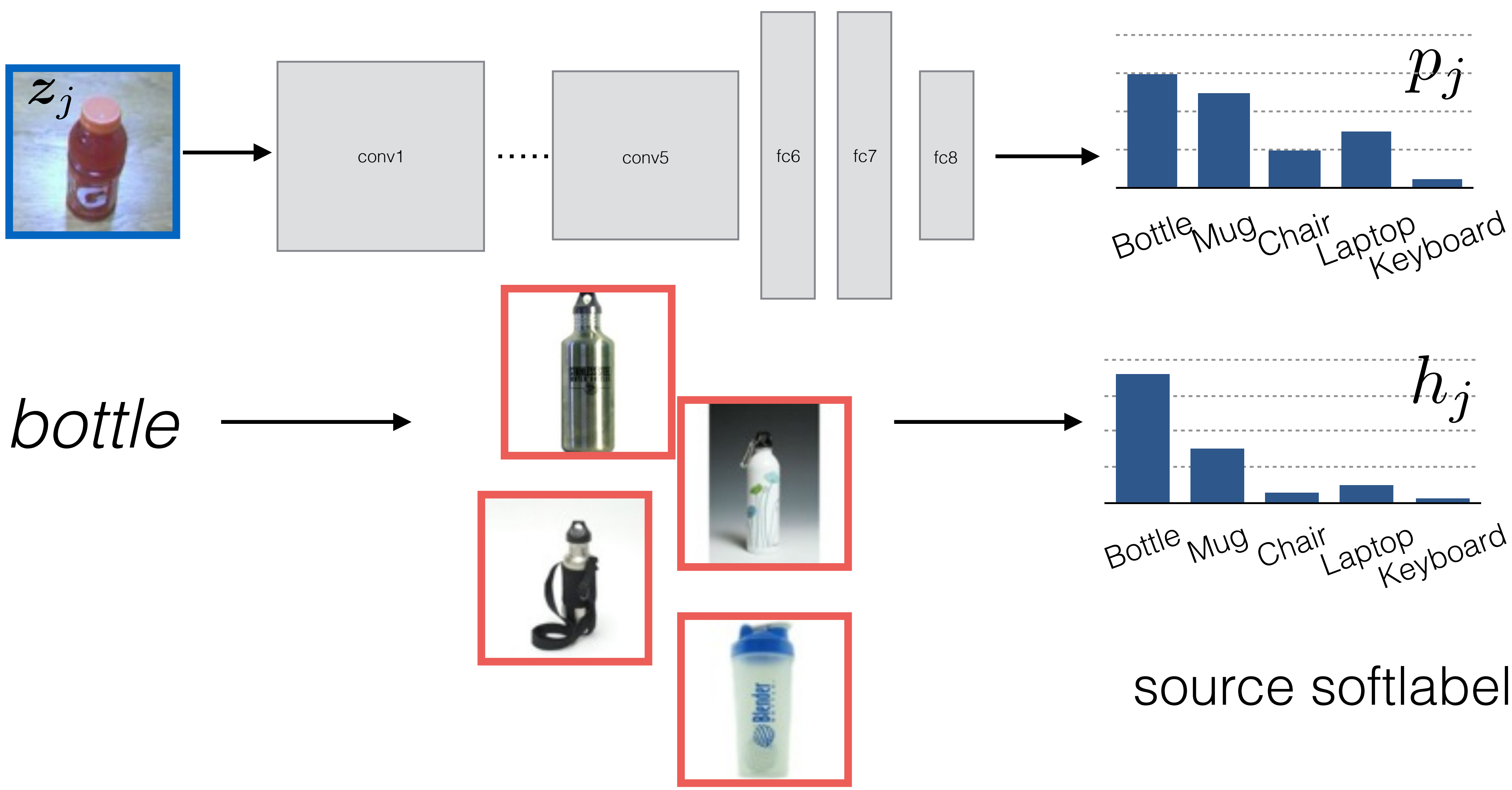
$$H(h, p) = E_h[-\log p]$$



[ICCV 2015]

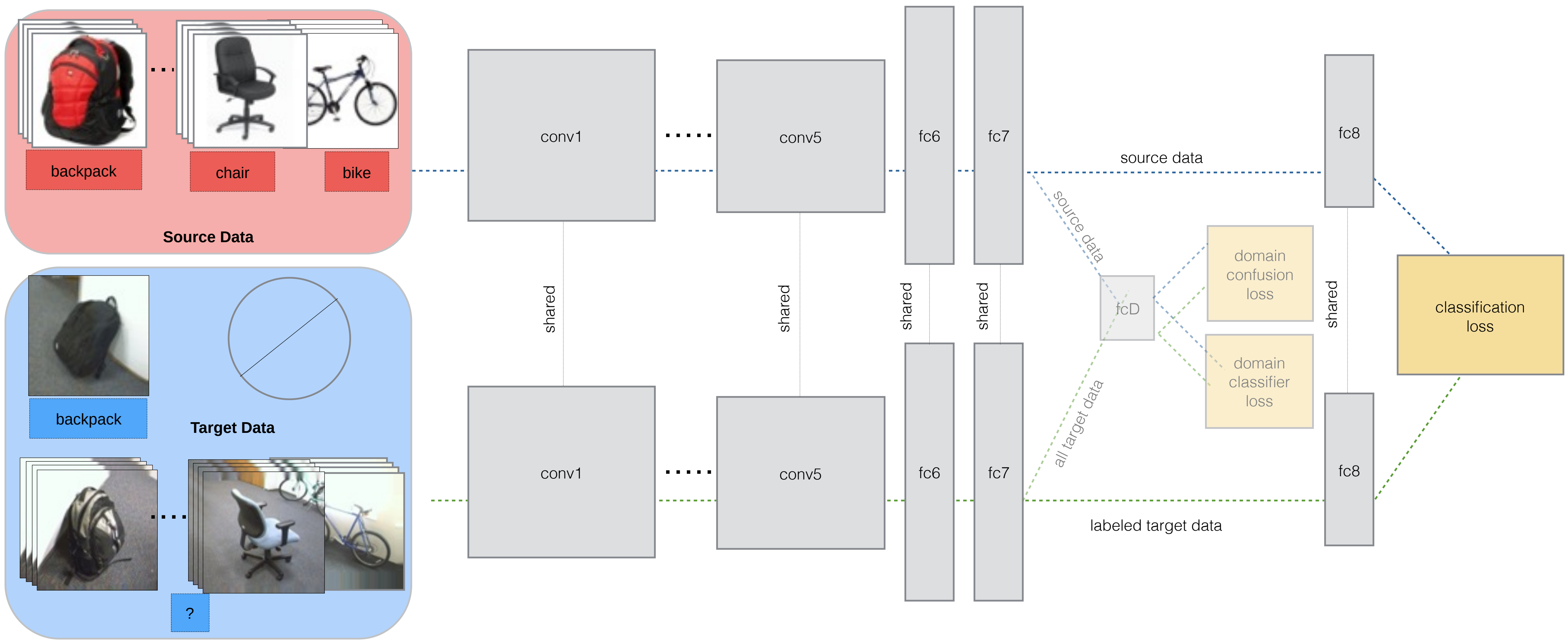
Class correlation transfer loss

$$H(h, p) = E_h[-\log p]$$

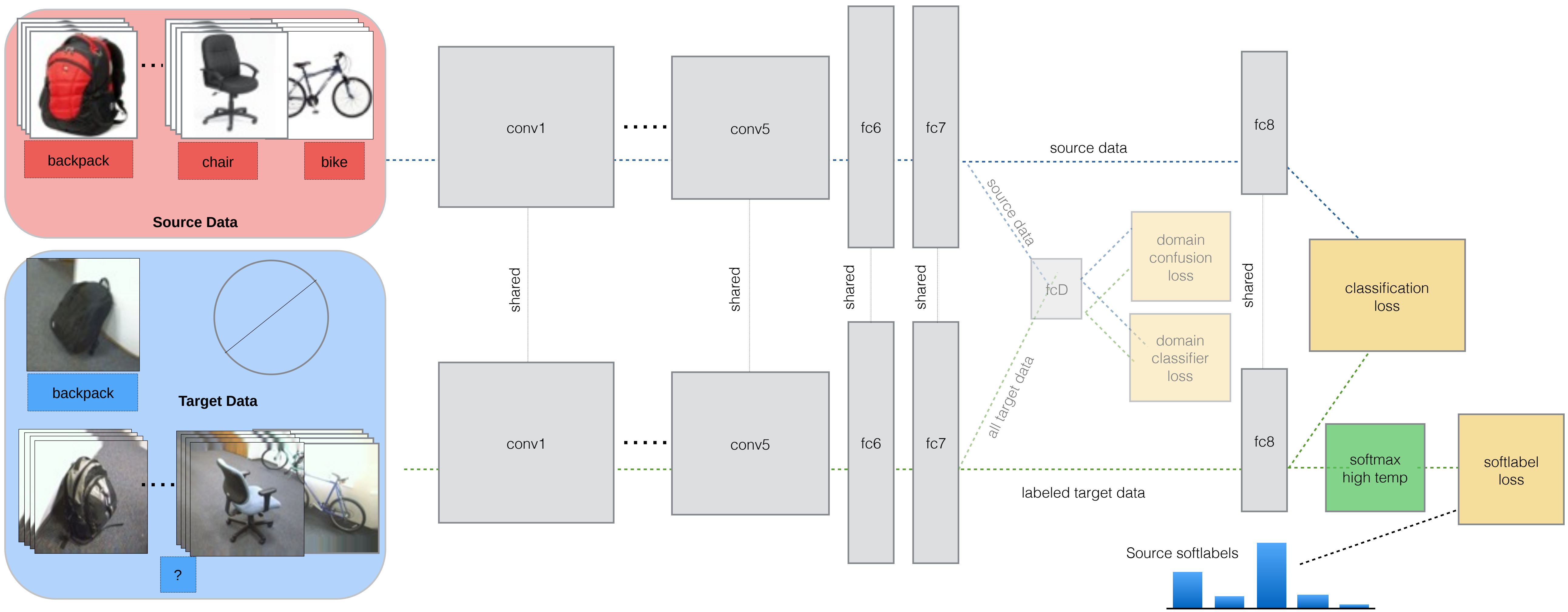


[ICCV 2015]

Class correlation transfer loss

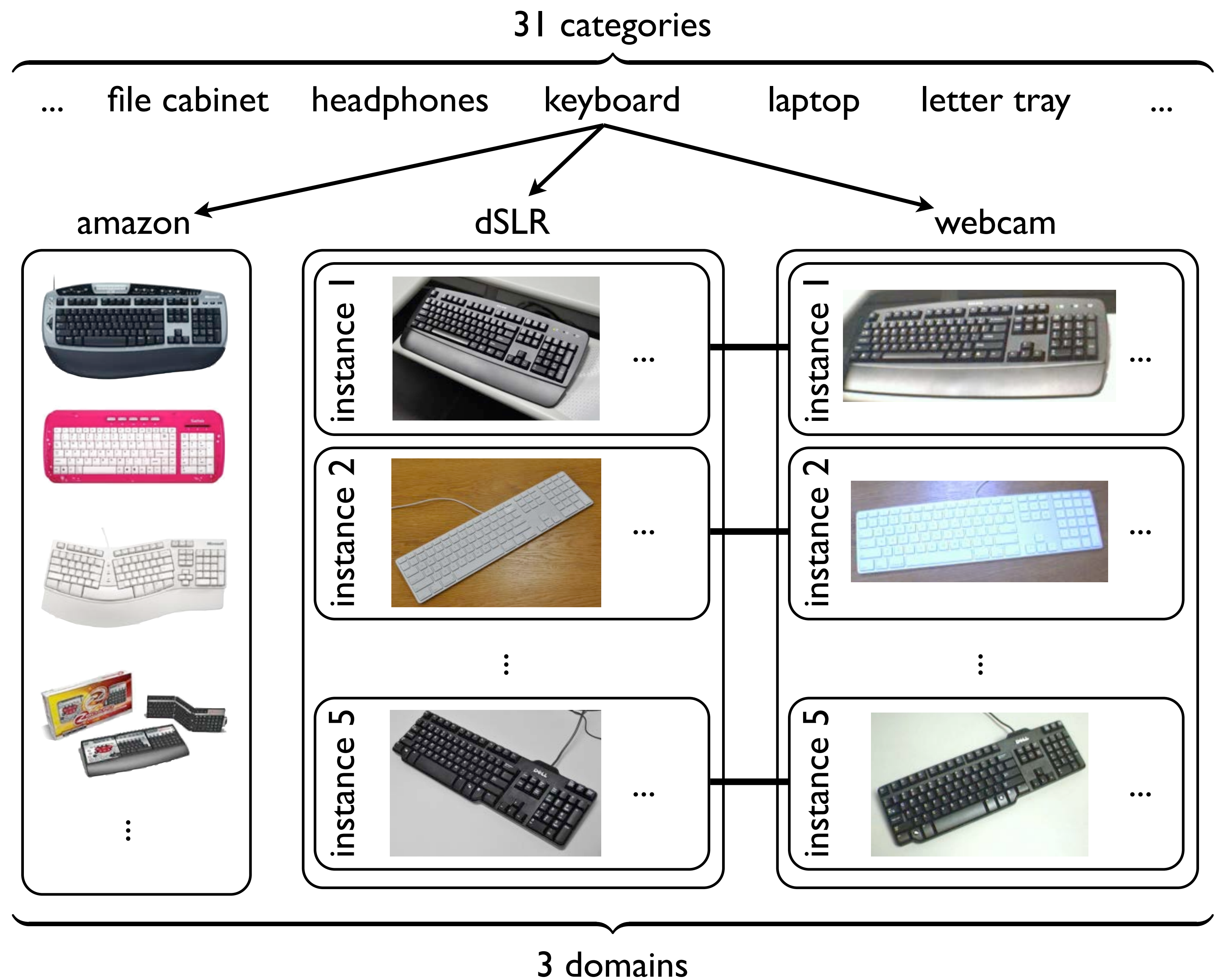


Class correlation transfer loss



Office dataset Experiment

- all classes have source labeled examples
- 15 classes have target labeled examples
- evaluate on remaining 16 classes

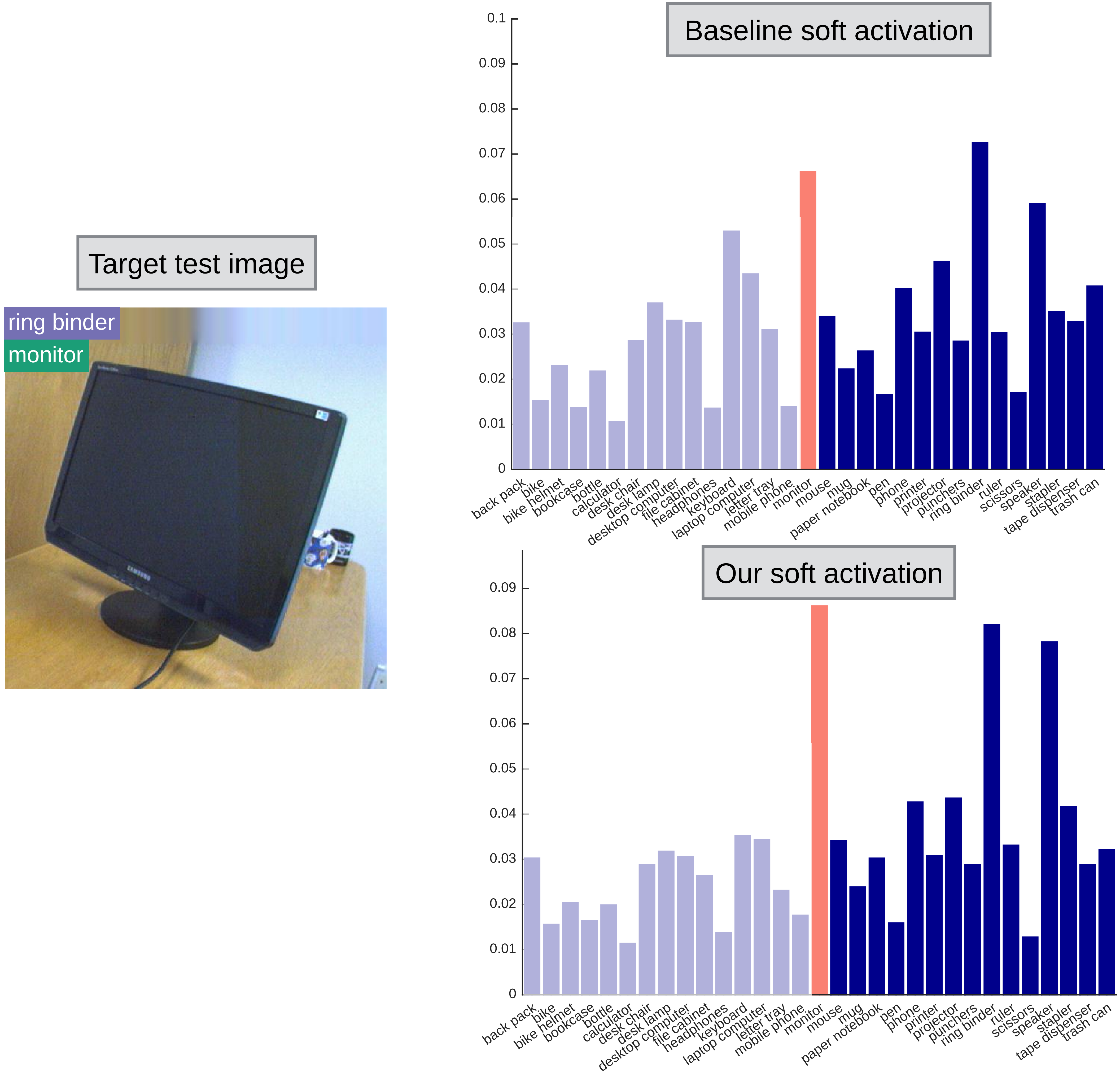


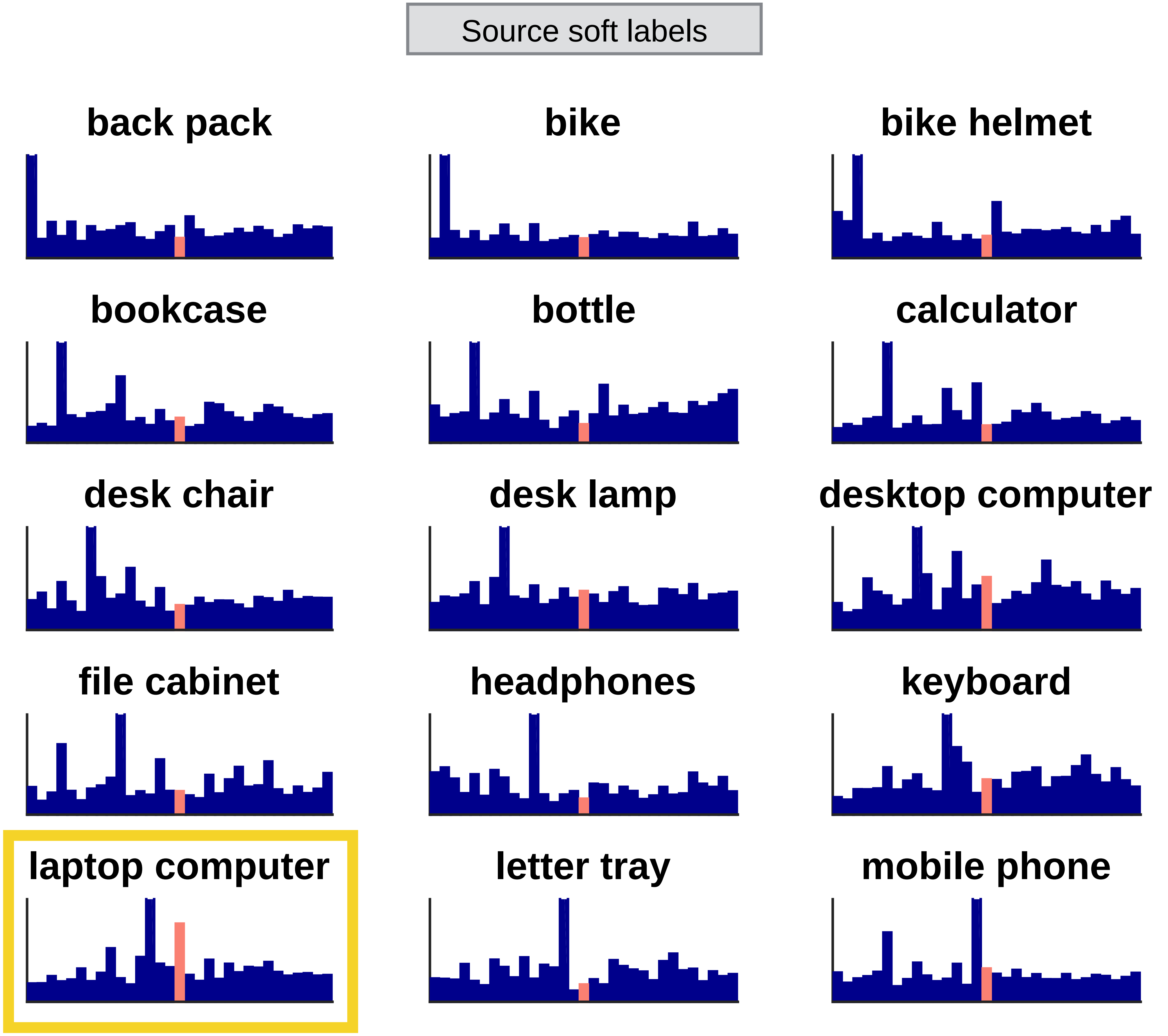
[saenko`10]

Office dataset Experiment

	$A \rightarrow W$	$A \rightarrow D$	$D \rightarrow A$	$D \rightarrow W$	$W \rightarrow A$	$W \rightarrow D$	Average
MMDT [18]	–	44.6 ± 0.3	–	–	–	58.3 ± 0.5	–
Source CNN	54.2 ± 0.6	63.2 ± 0.4	36.4 ± 0.1	89.3 ± 0.5	34.7 ± 0.1	94.5 ± 0.2	62.0
Ours: dom confusion only	55.2 ± 0.6	63.7 ± 0.9	41.2 ± 0.1	**91.3 ± 0.4**	**41.1 ± 0.0**	96.5 ± 0.1	64.8
Ours: soft labels only	56.8 ± 0.4	65.2 ± 0.9	41.7 ± 0.3	89.6 ± 0.1	38.8 ± 0.4	96.5 ± 0.2	64.8
Ours: dom confusion+soft labels	**59.3 ±0.6**	**68.0±0.5**	**43.1± 0.2**	90.0± 0.2	40.5±0.2	**97.5± 0.1**	**66.4**

Multiclass accuracy over 16 classes which lack target labels


Target test image
ring binder
monitor
Baseline soft activation
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
back pack
bike
bike helmet
bookcase
bottle
calculator
desk chair
desk lamp
desktop computer
file cabinet
headphones
keyboard
laptop computer
letter tray
mobile phone
monitor
mouse
mug
paper notebook
pen
phone
printer
projector
punchers
ring binder
ruler
scissors
speaker
stapler
tape dispenser
trash can
Our soft activation
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
back pack
bike
bike helmet
bookcase
bottle
calculator
desk chair
desk lamp
desktop computer
file cabinet
headphones
keyboard
laptop computer
letter tray
mobile phone
monitor
mouse
mug
paper notebook
pen
phone
printer
projector
punchers
ring binder
ruler
scissors
speaker
stapler
tape dispenser
trash can


Source soft labels
back pack
bike
bike helmet
bookcase
bottle
calculator
desk chair
desk lamp
desktop computer
file cabinet
headphones
keyboard
laptop computer
letter tray
mobile phone

Cross-dataset Experiment Setup

Source: ImageNet

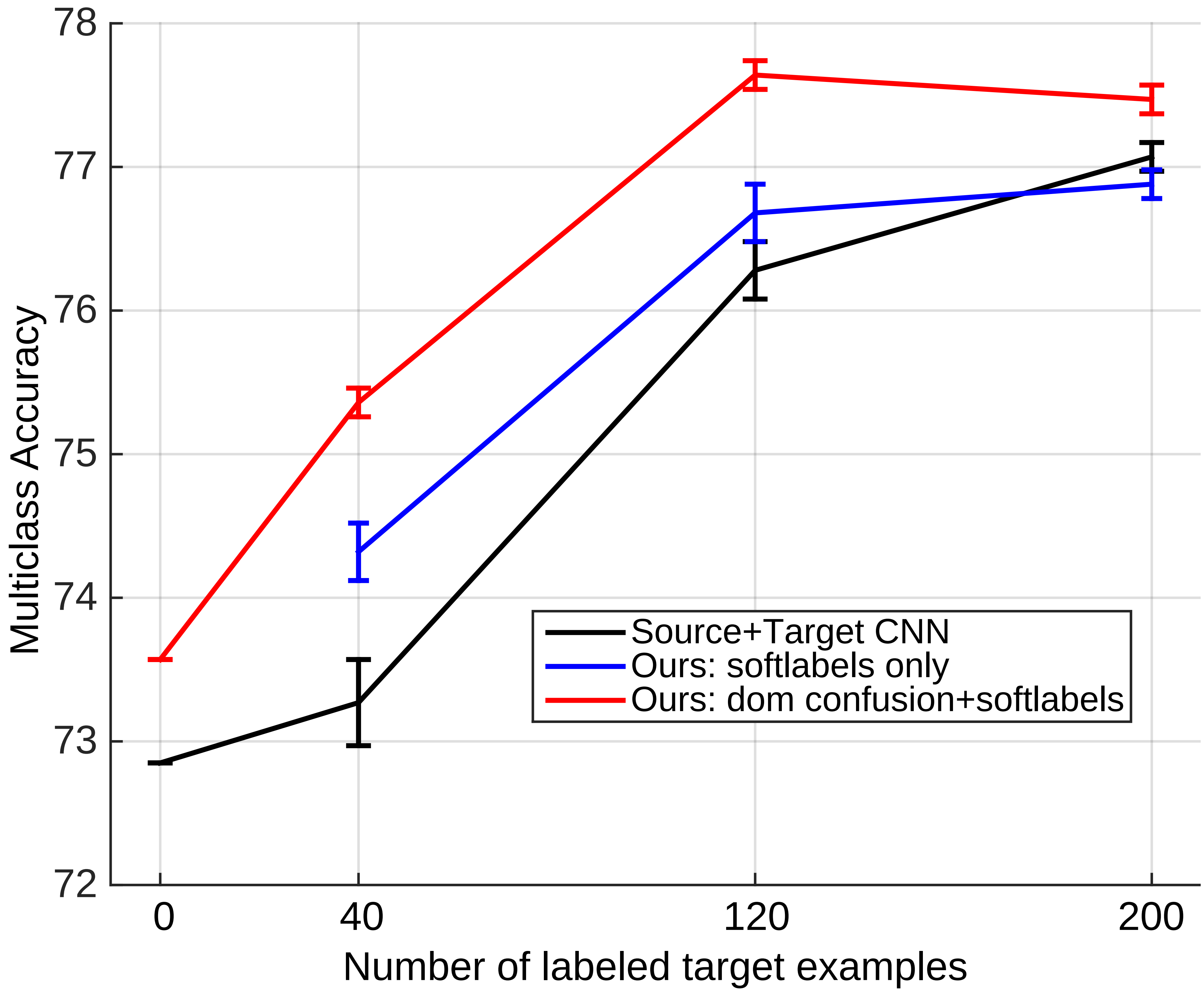
Target: Caltech256

40 categories

Evaluate adaptation performance with 0,1,3,5 target labeled examples per class

[tommasi`14]

ImageNet adapted to Caltech



[ICCV 2015]

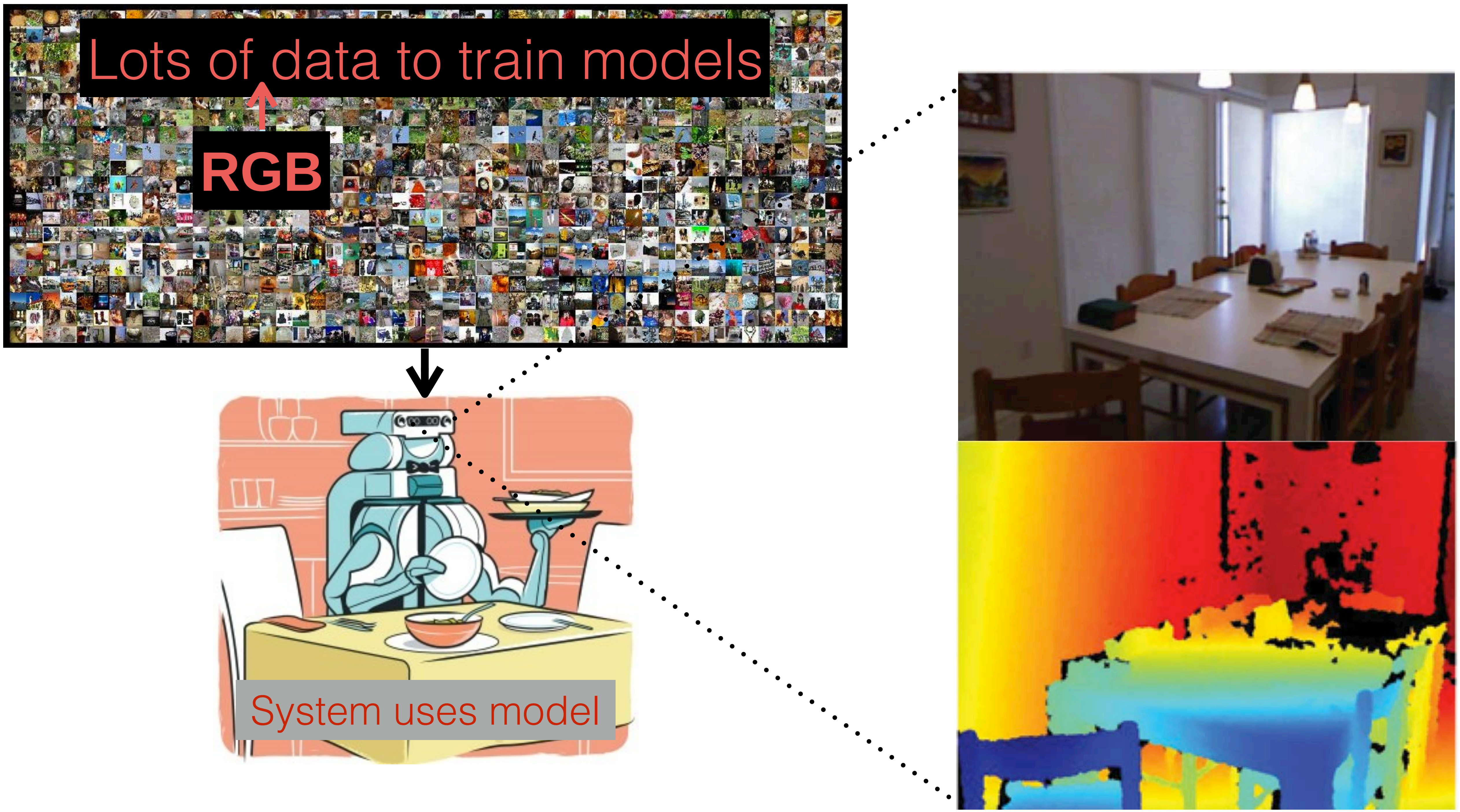
Summary: simultaneous transfer across domains and tasks

Domain confusion aligns the distributions

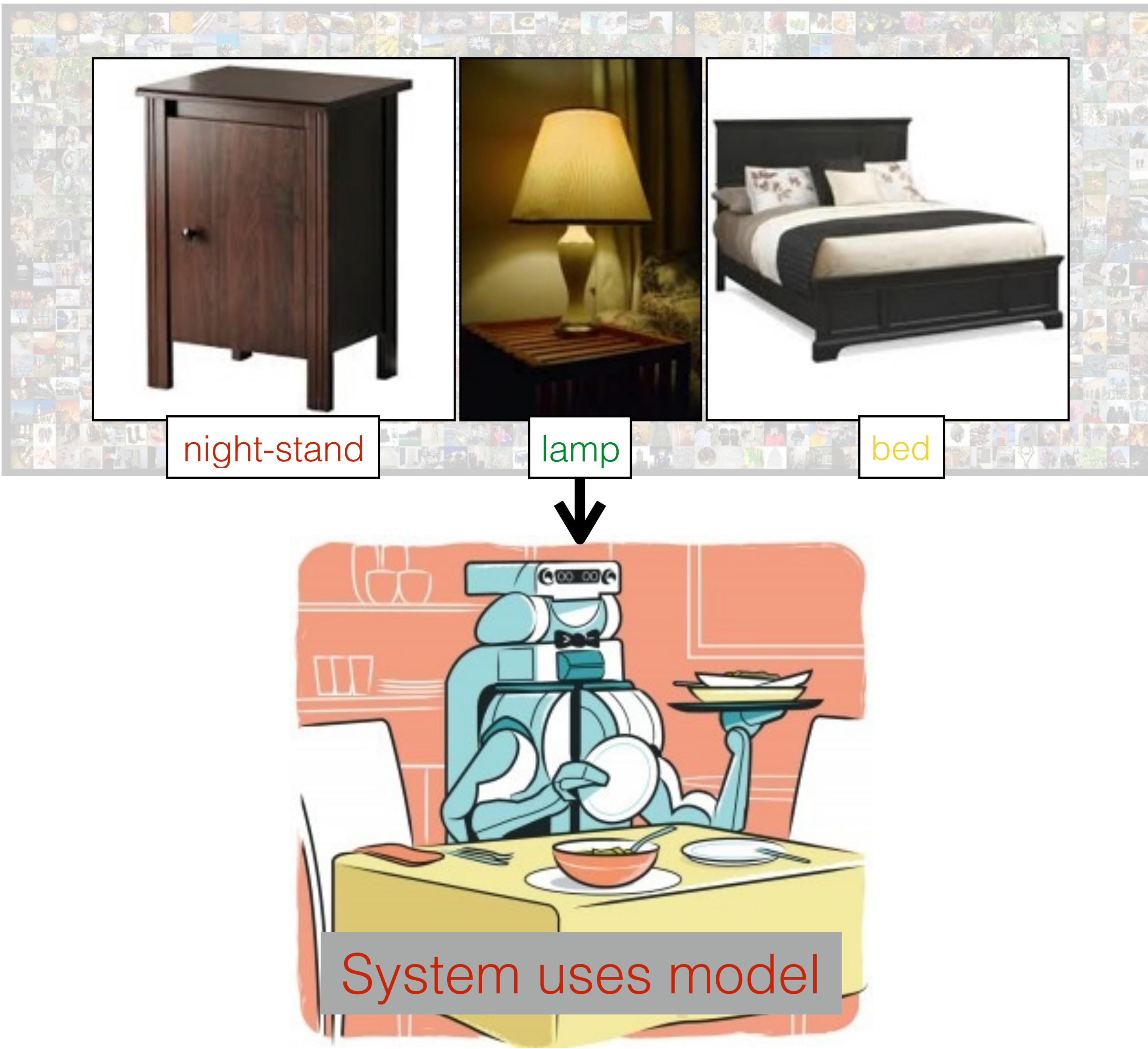
Softlabels transfer class correlations

Paper presented in poster session Wednesday 12/16 4B

Discrepancy due to modality shift



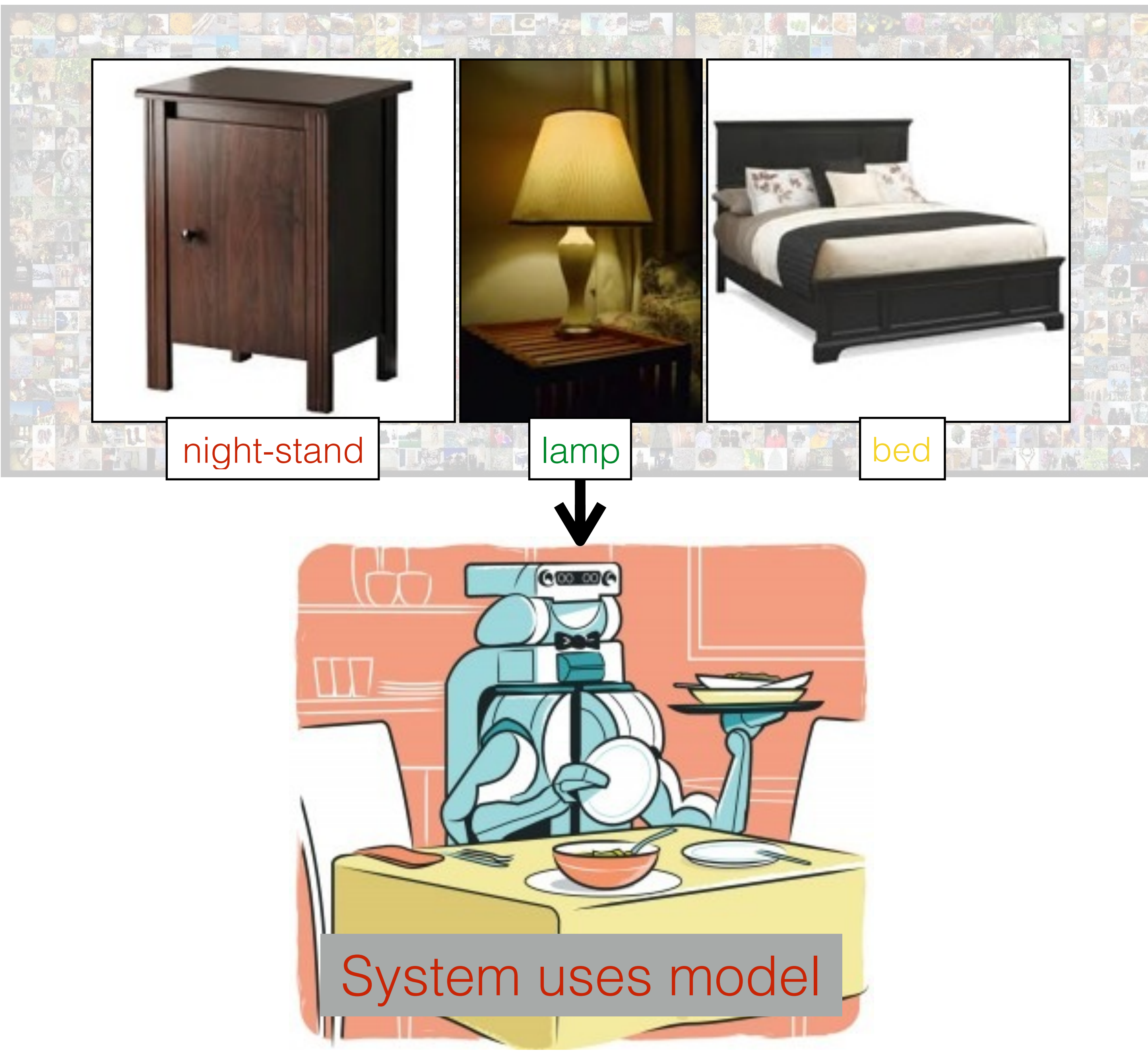
Label space discrepancy



Current output

[NIPS 2014, CVPR 2015]

Label space discrepancy

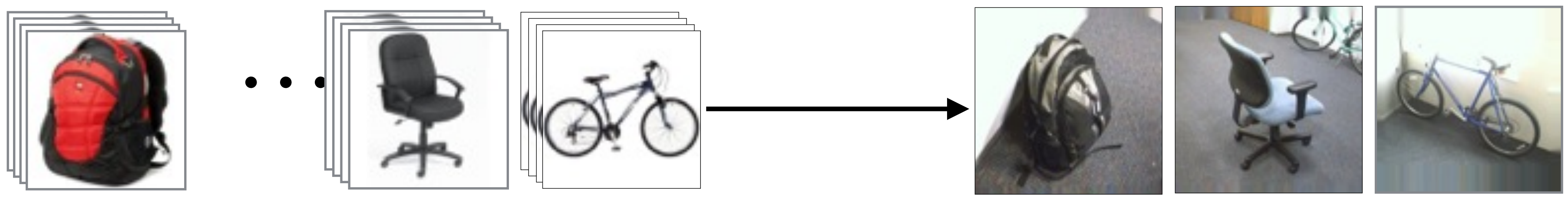


Desired output

[NIPS 2014, CVPR 2015]

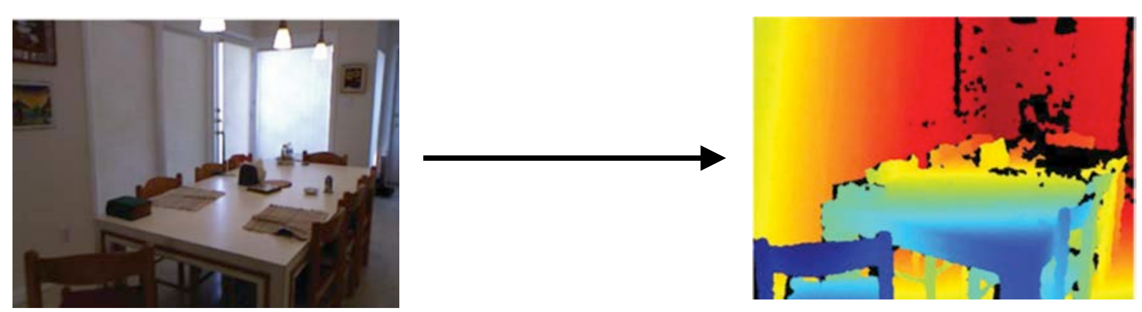
Adapting Deep Visual Models

Adapting across domains



Adapting across tasks

Adapting across modalities



Error bounds on adapted deep models

Generally applicable to adaptation with deep learning in AI

Thank you.