Efficient Learning of Domain Invariant Image Representations


Judy Hoffman


Erik Rodner


Jeff Donahue


Trevor Darrell


Kate Saenko

UC Berkeley; ICSI; University of Massachusetts, Lowell

ICLR May 2, 2013

What representation should we use for classifying backpacks?

What representation should we use for classifying backpacks?

What representation should we use for classifying backpacks?

A lot can change!


digital SLR


webcam

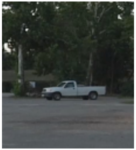

Close-up


Far-away

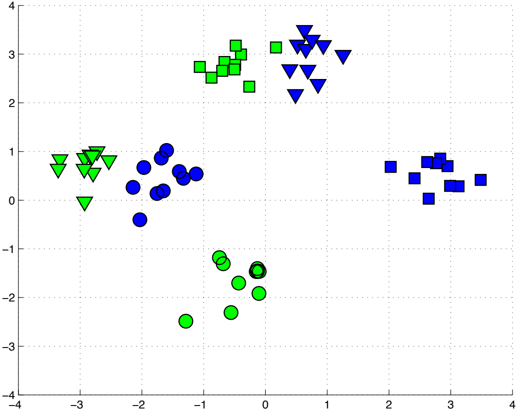

amazon.com


Consumer images


FLICKR

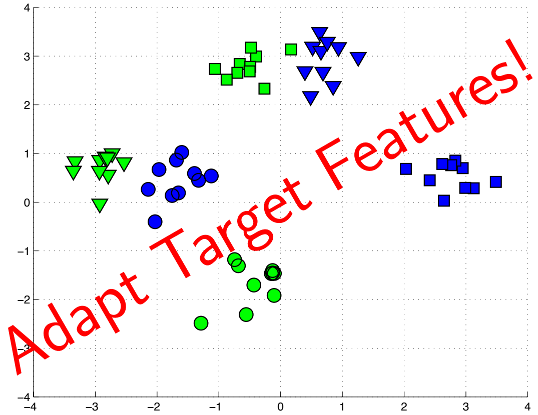

CCTV

Assume train/test data have different distributions



Source (train) Domain   Target (test) Domain

Assume train/test data have different distributions



Source (train) Domain Target (test) Domain

Previous Work: Domain Adaptation

- Feature Transformations: Saenko (ECCV '10), Kulis (CVPR '11), Duan (ICML '12)
- Manifold Distance: Gopalan (ICCV '11), Gong (CVPR '12), Gong (ICML '13)
- Parameter Adaptation: Yang (ACM MULTIMEDIA '07), Duan (CVPR '09), Bergamo (NIPS '10), Ayatar (ICCV '11)

Problem Statement: Semi-supervised Domain Adaptation

Given

- Labeled source data, $(X, Y) = \{(x_i, y_i)\}_{i=1}^{n_S}$
- Labeled target data, $(\tilde{X}, \tilde{Y}) = \{(\tilde{x}_j, \tilde{y}_j)\}_{j=1}^{m}, m << n_S$
  may not have target examples from all categories!

Problem Statement: Semi-supervised Domain Adaptation

Given

- Labeled source data, $(X, Y) = \{(x_i, y_i)\}_{i=1}^{n_S}$
- Labeled target data, $(\tilde{X}, \tilde{Y}) = \{(\tilde{x}_j, \tilde{y}_j)\}_{j=1}^{m}, m << n_S$
  may not have target examples from all categories!

Goal: Use category level constraints between domains to learn a new target representation.

Problem Statement: Semi-supervised Domain Adaptation

Given

- Labeled source data, $(X, Y) = \{(x_i, y_i)\}_{i=1}^{n_S}$
- Labeled target data, $(\tilde{X}, \tilde{Y}) = \{(\tilde{x}_j, \tilde{y}_j)\}_{j=1}^{m}, m << n_S$
  may not have target examples from all categories!

Goal: Use category level constraints between domains to learn a new target representation.

Our Method: Combine transformation, W, learning with parameter adaptation.

Problem Statement: Semi-supervised Domain Adaptation

Given

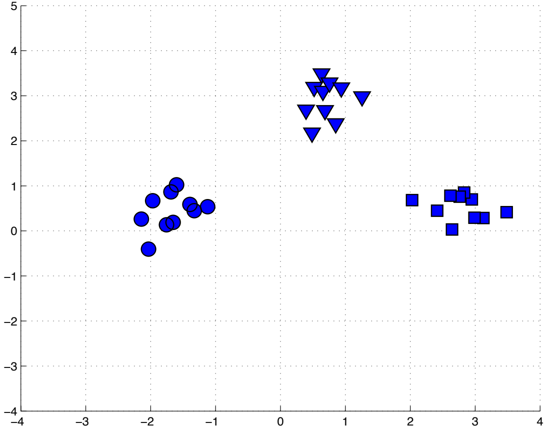
- Labeled source data, $(X, Y) = \{(x_i, y_i)\}_{i=1}^{n_S}$
- Labeled target data, $(\tilde{X}, \tilde{Y}) = \{(\tilde{x}_j, \tilde{y}_j)\}_{j=1}^{m}, m << n_S$
  may not have target examples from all categories!

Goal: Use category level constraints between domains to learn a new target representation.

Our Method: Combine transformation, W, learning with parameter adaptation.

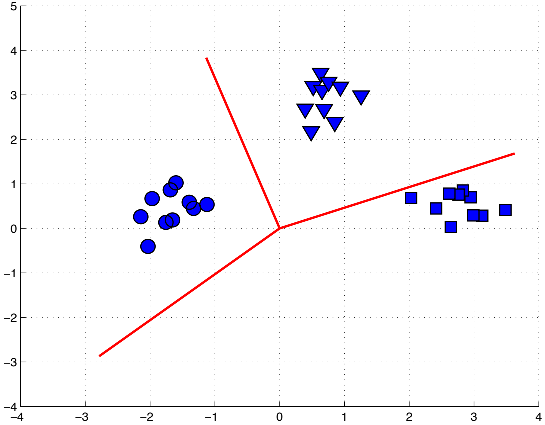
New target representation: $\tilde{X}_{\text{new}} = W\tilde{X}$

Desired Transformation Properties



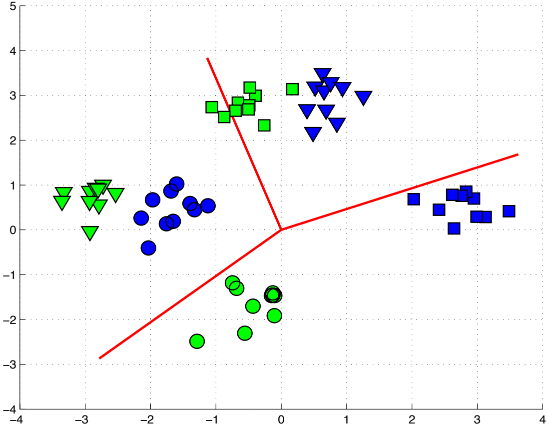
Desired Transformation Properties

1. Optimize Classification Objective



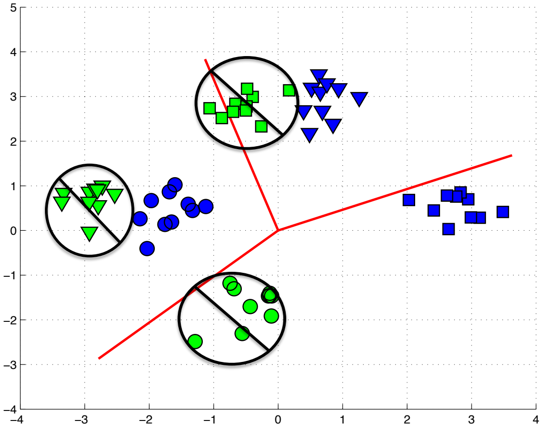
Desired Transformation Properties

1. Optimize Classification Objective



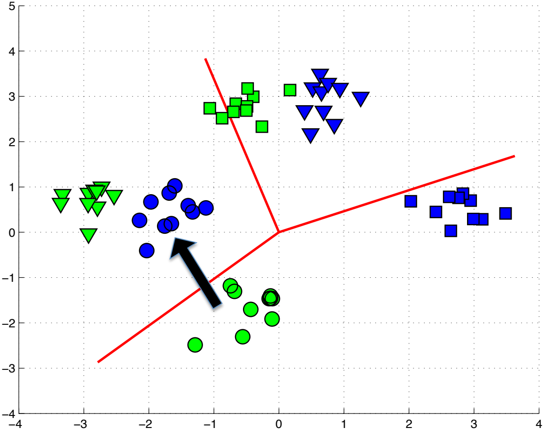
Desired Transformation Properties

1. Optimize Classification Objective



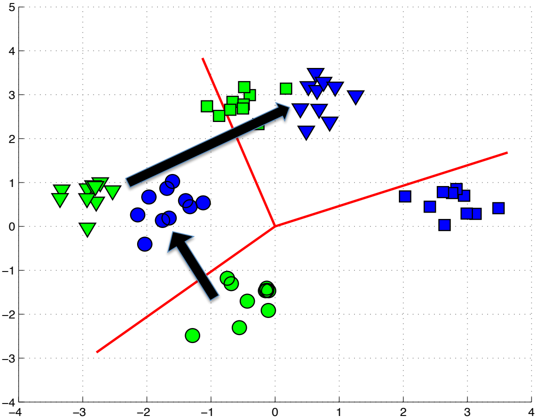
Desired Transformation Properties

1. Optimize Classification Objective



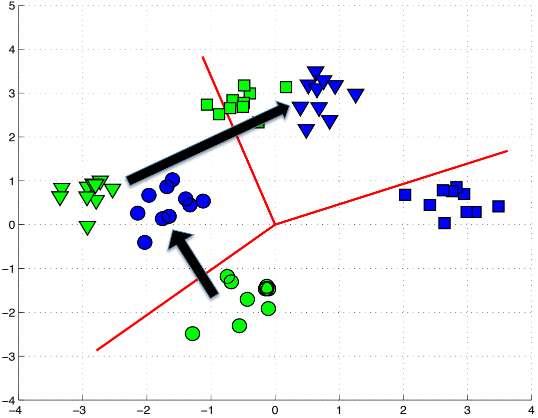
Desired Transformation Properties

1. Optimize Classification Objective



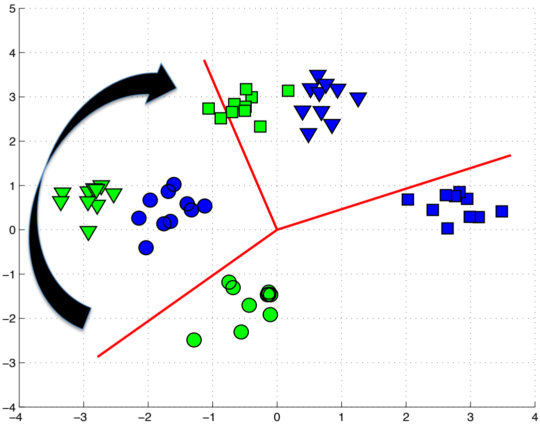
Desired Transformation Properties

1. Optimize Classification Objective
2. Category Invariant



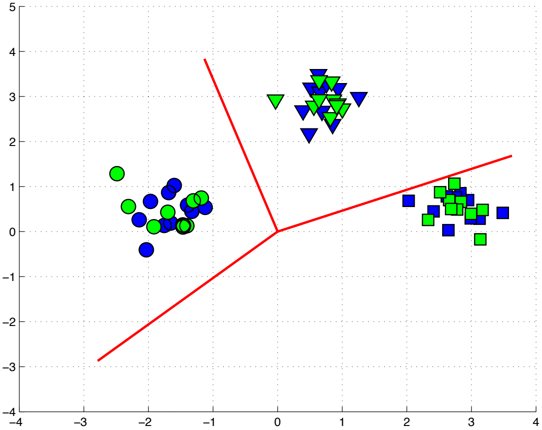
Desired Transformation Properties

1. Optimize Classification Objective
2. Category Invariant



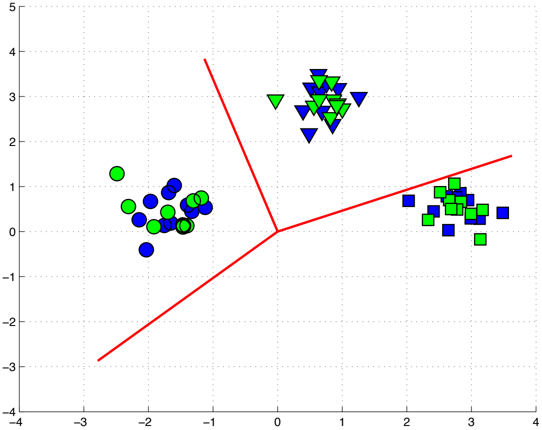
Desired Transformation Properties

1. Optimize Classification Objective
2. Category Invariant



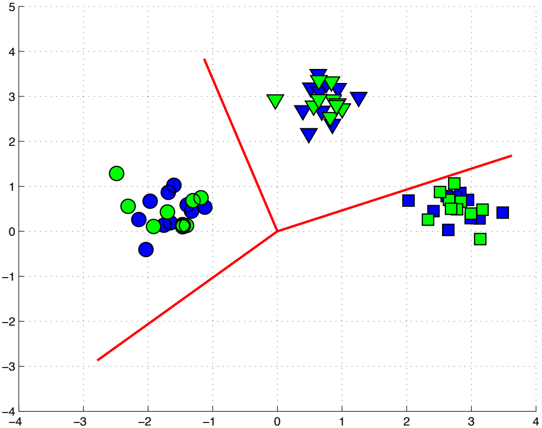
Desired Transformation Properties

1. Optimize Classification Objective
2. Category Invariant
3. Learned Efficiently



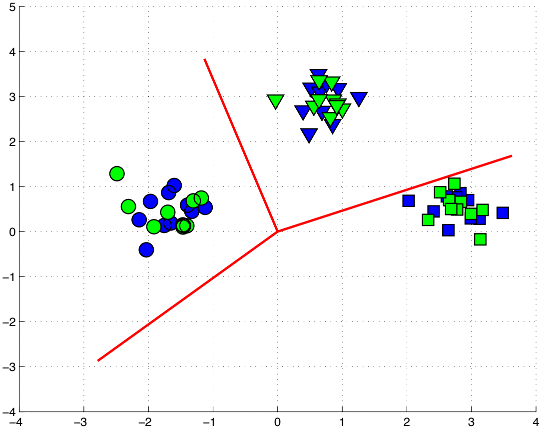
Desired Transformation Properties

1. Optimize Classification Objective
2. Category Invariant
3. Learned Efficiently
4. Allow heterogeneous input feature spaces
   - S: $\boxed{}$ T: $\boxed{}$



Desired Transformation Properties

1. Optimize Classification Objective
2. Category Invariant
3. Learned Efficiently
4. Allow heterogeneous input feature spaces
   - S:  T: 



In the setting of heterogeneous input feature spaces we are able to boost accuracy up to 20%

Joint Objective

For a K category problem, with $\tilde{K} \leq K$ target categories labeled. Choose **parameters**: γ, C, $\tilde{C}$ and solve:

$$\begin{aligned} \min_{W,\theta,\eta,\tilde{\eta}} \quad & \frac{\gamma}{2}\|W\|_F^2 + \frac{1}{2}\sum_{k=1}^{K}\|\theta_k\|_2^2 + C\sum_{i=1,k=1}^{n_S,K}\eta_{ik} + \tilde{C}\sum_{j=1,k=1}^{m,\tilde{K}}\eta_{jk} \\ \text{s.t} \quad & y_{ik}\theta_k^T x_i \geq 1 - \eta_{ik} \\ & \tilde{y}_{jk}\theta_k^T W\tilde{x}_j \geq 1 - \tilde{\eta}_{jk} \\ & \eta_{ik} \geq 0, \quad \eta_{jk} \geq 0 \end{aligned}$$

Joint Objective

For a K category problem, with $\tilde{K} \leq K$ target categories labeled. Choose **parameters**: $\gamma, C, \tilde{C}$ and solve:

$$\begin{aligned} \min_{W,\theta,\eta,\tilde{\eta}} \quad & \frac{\gamma}{2}\|W\|_F^2 + \frac{1}{2}\sum_{k=1}^{K}\|\theta_k\|_2^2 + C\sum_{i=1,k=1}^{n_S,K}\eta_{ik} + \tilde{C}\sum_{j=1,k=1}^{m,\tilde{K}}\eta_{jk} \\ \text{s.t} \quad & y_{ik}\theta_k^T x_i \geq 1 - \eta_{ik} \\ & \tilde{y}_{jk}\theta_k^T W\tilde{x}_j \geq 1 - \tilde{\eta}_{jk} \\ & \eta_{ik} \geq 0, \quad \eta_{jk} \geq 0 \end{aligned}$$

Alternating Minimization

Alternating Minimization

1. Given previous W and chosen $C, \tilde{C}$.

Alternating Minimization

1. Given previous W and chosen $C, \tilde{C}$.

$$\begin{aligned}
\min_{\theta,\eta,\tilde{\eta}} \quad & \frac{1}{2}\sum_k \|\theta_k\|_2^2 + C\sum_{i=1,k=1}^{n_S,K}\eta_{ik} + \tilde{C}\sum_{j=1,k=1}^{m,\tilde{K}}\eta_{jk} \\
\text{s.t} \quad & y_{i,k}\theta_k^T x_i \geq 1-\eta_{ik} \\
& \tilde{y}_{j,k}\theta_k^T W\tilde{x}_j \geq 1-\tilde{\eta}_{jk} \\
& \eta_{ik}\geq 0, \quad \eta_{jk}\geq 0
\end{aligned}$$

Alternating Minimization

1. Given previous W and chosen $C, \tilde{C}$.

$$\begin{aligned} \min_{\theta,\eta,\tilde{\eta}} \quad & \frac{1}{2}\sum_k \|\theta_k\|_2^2 + C\sum_{i=1,k=1}^{n_S,K} \eta_{ik} + \tilde{C}\sum_{j=1,k=1}^{m,\tilde{K}} \eta_{jk} \\ \text{s.t} \quad & y_{i,k}\theta_k^T x_i \geq 1 - \eta_{ik} \\ & \tilde{y}_{j,k}\theta_k^T W\tilde{x}_j \geq 1 - \tilde{\eta}_{jk} \\ & \eta_{ik} \geq 0, \quad \eta_{jk} \geq 0 \end{aligned}$$

2. Given previous θ_k and chosen $\gamma, \tilde{C}$

Alternating Minimization

1. Given previous W and chosen $C, \tilde{C}$.

$$\begin{aligned} \min_{\theta,\eta,\tilde{\eta}} \quad & \frac{1}{2}\sum_k \|\theta_k\|_2^2 + C\sum_{i=1,k=1}^{n_S,K}\eta_{ik} + \tilde{C}\sum_{j=1,k=1}^{m,\tilde{K}}\eta_{jk} \\ \text{s.t} \quad & y_{i,k}\theta_k^T x_i \geq 1-\eta_{ik} \\ & \tilde{y}_{j,k}\theta_k^T W\tilde{x}_j \geq 1-\tilde{\eta}_{jk} \\ & \eta_{ik}\geq 0, \quad \eta_{jk}\geq 0 \end{aligned}$$

2. Given previous θ_k and chosen $\gamma, \tilde{C}$

$$\begin{aligned} \min_{W,\tilde{\eta}} \quad & \frac{\gamma}{2}\|W\|_F^2 + \tilde{C}\sum_{j,k}\tilde{\eta}_{jk} \\ \text{s.t} \quad & \tilde{y}_{j,k}\theta_k^T W\tilde{x}_j \geq 1-\tilde{\eta}_{jk} \\ & \tilde{\eta}_{jk}\geq 0 \end{aligned}$$

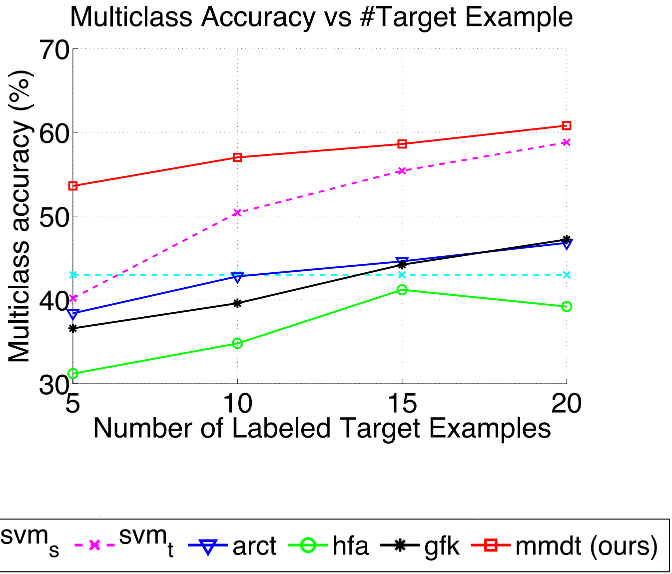
Experimental Setup: *Bing-Caltech256*

Source: Bing

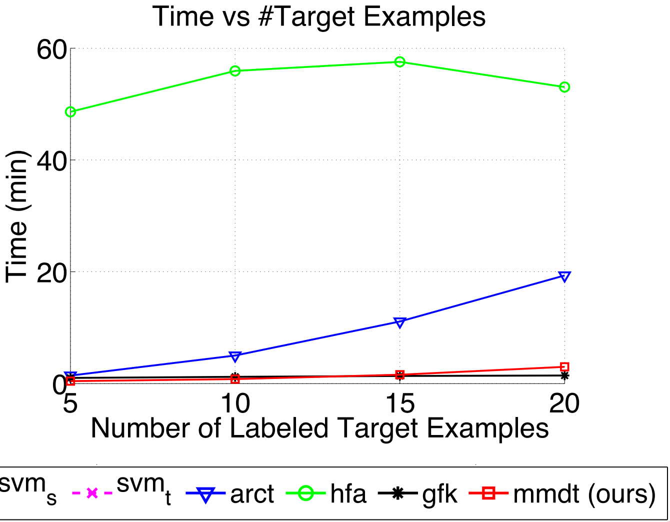
- 20 first categories of *Caltech256*
- Classeme features*
- 50 labeled examples per category

*L. Torresani, M. Szummer, A. Fitzgibbon. Efficient Object Category Recognition Using Classemes. ECCV, 2010.

Results *Bing-Caltech256*: Semi-supervised DA



Results *Bing-Caltech256*: Semi-supervised DA



Experimental Setup: *Office*

Amazon

Webcam

Dslr

- SURF BoW 800 dimension features
- 31 categories commonly found in an office

Results *Office*: Asymmetric & Novel Test Categories

20, 8, and 3 labeled examples per category in `Amazon`, `Webcam`, and `Dslr`, respectively.

Results *Office*: Asymmetric & Novel Test Categories

20, 8, and 3 labeled examples per category in `Amazon`, `Webcam`, and `Dslr`, respectively.

source	**svm**$_t$	**arc-t**	**hfa**	**mmdt**
`amazon`	52.9 ± 0.7	58.2 ± 0.6	57.8 ± 0.6	62.3 ± 0.8
`webcam`	51.8 ± 0.6	58.2 ± 0.7	60.0 ± 0.6	63.3 ± 0.5

Results for asymmetric transform: source (800 dimension) to target `dslr-600`.

Results *Office*: Asymmetric & Novel Test Categories

20, 8, and 3 labeled examples per category in `Amazon`, `Webcam`, and `Dslr`, respectively.

source	**svm$_t$**	**arc-t**	**hfa**	**mmdt**
`amazon`	52.9 ± 0.7	58.2 ± 0.6	57.8 ± 0.6	62.3 ± 0.8
`webcam`	51.8 ± 0.6	58.2 ± 0.7	60.0 ± 0.6	63.3 ± 0.5

Results for asymmetric transform: source (800 dimension) to target `dslr-600`.

source	**svm$_s$**	**arc-t**	**gfk**	**mmdt**
`amazon`	10.3 ± 0.6	41.4 ± 0.3	38.9 ± 0.4	44.6 ± 0.3
`webcam`	51.6 ± 0.5	59.4 ± 0.4	62.9 ± 0.5	58.3 ± 0.5

Results for novel test categories: source to target `dslr`.

Summary

- We advocate a representation learning perspective for domain adaptation.

Summary

- We advocate a representation learning perspective for domain adaptation.
- We learn supervised feature representations for classification across domains.
  1. Transferred to novel categories
  2. Allowed different input feature spaces
  3. Are fast to learn
  4. Higher recognition performance over baselines

Summary

- We advocate a representation learning perspective for domain adaptation.
- We learn supervised feature representations for classification across domains.
  1. Transferred to novel categories
  2. Allowed different input feature spaces
  3. Are fast to learn
  4. Higher recognition performance over baselines

- **Next:**
  - Adapting different modalities
  - Nonlinear transformations – Deep Learning

Thank you!