Enlarging Discriminative Power by Adding an Extra Class in Unsupervised Domain Adaptation

Hai H. Tran, Sumyong Ahn, Taeyoung Lee, and Yung Yi

Introduction

Domain Adaptation
- Source data distribution ≠ target data distribution, but same task
- Goal: learn a well-performing model on target data

Unsupervised Domain Adaptation
- Source data: all samples are labeled
- Target data: no labels at all

Background

Key Requirements for a Successful Adaptation
- Domain-invariance: be consistent in different domains
- Discriminativeness: be easy to classify

Evaluation

Domain Adaptation Tasks

<table>
<thead>
<tr>
<th>Source Target</th>
<th>MNIST</th>
<th>SVHN</th>
<th>SVHN</th>
<th>MNIST</th>
<th>DIGITS</th>
<th>SVHN</th>
<th>CIFAR</th>
<th>STL</th>
<th>CIFAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>DANN[1]</td>
<td>35.7</td>
<td>71.1</td>
<td>81.5</td>
<td>90.3</td>
<td></td>
<td>90.3</td>
<td>77.6</td>
<td>62.1</td>
<td>61.5</td>
</tr>
<tr>
<td>DANN[10]</td>
<td>82.7</td>
<td>83.2</td>
<td>91.2</td>
<td>92.7</td>
<td></td>
<td>92.7</td>
<td>94.6</td>
<td>71.2</td>
<td>65.1</td>
</tr>
<tr>
<td>ATT[9]</td>
<td>51.1</td>
<td>89.1</td>
<td>92.4</td>
<td>94.2</td>
<td></td>
<td>94.2</td>
<td>94.2</td>
<td>92.4</td>
<td>92.4</td>
</tr>
<tr>
<td>MCD[20]</td>
<td>96.2</td>
<td></td>
<td></td>
<td>96.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>99.9</td>
</tr>
<tr>
<td>DA Associate[7]</td>
<td>97.6</td>
<td>98.5</td>
<td></td>
<td>91.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>99.9</td>
</tr>
<tr>
<td>GAG[20]</td>
<td>74.6</td>
<td>96.7</td>
<td>94.9</td>
<td>93.1</td>
<td>93.1</td>
<td>93.1</td>
<td>77.0</td>
<td>61.5</td>
<td></td>
</tr>
</tbody>
</table>

Without instance-normalized input

- VADA[14]: 47.5, 97.9, 100.7, 94.8, 80.0, 73.5
- CoDA[13]: 55.3, 98.8, 99.0, 96.1, 81.4, 76.4
- GADA* (ours): 69.9, 98.2, 98.2, 94.9, 79.7, 73.0
- GADA (ours): 72.2, 98.9, 99.1, 95.7, 80.5, 75.1
- VADA+DIRT-T[14]: 54.5, 99.4, 98.9, 96.1, 75.3
- CoDA+DIRT-T[13]: 63.0, 99.4, 99.1, 96.5, 77.6
- GADA*+DIRT-T (ours): 82.4, 99.4, 98.9, 96.3, 75.3
- GADA+DIRT-T (ours): 83.7, 99.6, 99.3, 96.6, 76.5

With instance-normalized input

- VADA[14]: 73.3, 94.5, 95.7, 94.9, 78.3, 71.4
- CoDA[13]: 81.7, 98.7, 98.0, 96.0, 80.6, 74.7
- GADA* (ours): 78.7, 99.0, 97.2, 95.3, 78.9, 72.1
- GADA (ours): 83.6, 98.8, 98.2, 96.1, 80.1, 74.9
- VADA+DIRT-T[14]: 76.5, 99.4, 98.7, 96.2, 73.3
- CoDA+DIRT-T[13]: 88.0, 99.4, 98.8, 96.5, 75.9
- GADA*+DIRT-T (ours): 84.5, 99.4, 98.8, 96.5, 75.9
- GADA+DIRT-T (ours): 90.0, 99.6, 99.0, 96.8, 76.2

Method

Domain Discriminator

Input Space

Feature Space

OOC samples and (K + 1)th class increase the distance between real clusters, which helps the classifier place the decision boundary easier.

Key Idea

Improving discriminativeness by generating out-of-class samples, and classifying them to an extra (K + 1)th class.

Ablation study

<table>
<thead>
<tr>
<th>GADA Training Algorithm: alternative training with 3 steps</th>
<th>Original MNIST</th>
<th>Generated MNIST</th>
<th>Original SVHN</th>
<th>Generated SVHN</th>
</tr>
</thead>
<tbody>
<tr>
<td>max ( \min_{\theta_D} \mathcal{L}_0(\theta_D; D) + \lambda_0 \mathcal{L}_0(\theta_D, \theta_T; X_T, X_T) )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
</tr>
<tr>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
</tr>
<tr>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
</tr>
<tr>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
</tr>
<tr>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
</tr>
<tr>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
</tr>
</tbody>
</table>

GADA Training Algorithm:
- (a) achieves classification power from source data
- (b) improves domain-invariance
- (c) improves discriminative extraction power

Evaluation

Generated samples

Original MNIST
Generated MNIST
Original SVHN
Generated SVHN

Feature space comparison in MNIST → SVHN