Deep Transfer Learning with Joint Adaptation Networks
Mingsheng Long¹, Han Zhu¹, Jianmin Wang¹, and Michael I. Jordan²

¹School of Software, Tsinghua University, China  ²Department of EECS, UC Berkeley, USA

Summary
- A joint adaptation network framework for deep transfer learning
- Two main contributions:
  - Joint adaptation of multilayer features and classifier predictions
  - Adversarial adaptation with semi-parametric domain discriminator
- State-of-the-art results on visual & simulation-to-real datasets
- Open Problems
  - Randomized method for the multilinear operation across feature maps
  - Kernel approximation of the universal kernel for distribution matching
- Code: https://github.com/thuml/transfer-caffe

Deep Transfer Learning
- Deep learning across domains of different distributions \( P \neq Q \)

Deep Transfer Learning: Why?

1. Training Error high?
   - Bias
   - Optimal Bias Rate
   - Deeper Model
   - No

2. Train-Dev Error high?
   - Variance
   - Bigger Data
   - Regularization
   - No

3. Dev Error high?
   - Dataset Shift
   - Transfer Learning
   - Data Generation
   - No

4. Test Error high?
   - Overfit Dev Set
   - Bigger Dev Data
   - No

Main Idea of This Work
- Directly model and match joint distributions \( P(x, y) \) & \( Q(x, y) \)

Joint Adaptation Network (JAN)

Joint adaptation: match joint distributions of features/predictions

\[
\min_{\theta} \sum_{i=1}^{n} J(f(x_i^s), y_i^t) + \lambda \hat{D}_C(P, Q; \theta) 
\]

Joint Maximum Mean Discrepancy (JMMDD)

Distance between embeddings \( P(Z_1^t, \ldots, Z_L^t) \) & \( Q(Z_1^t, \ldots, Z_L^t) \)

\[
\hat{D}_C(P, Q; \theta) = \frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{L} k^t(z_i^l, z_l^t) 
\]

Optimal matching: maximize JMMDD as semi-parametric adversary

Experimental Results

<table>
<thead>
<tr>
<th>Method</th>
<th>A → W</th>
<th>D → W</th>
<th>D → A</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet</td>
<td>61.6 ± 0.5</td>
<td>96.4 ± 0.3</td>
<td>99.3 ± 0.2</td>
<td>98.0 ± 0.4</td>
</tr>
<tr>
<td>ResNet</td>
<td>70.9 ± 0.6</td>
<td>96.0 ± 0.3</td>
<td>98.9 ± 0.2</td>
<td>97.0 ± 0.4</td>
</tr>
<tr>
<td>JAN</td>
<td>74.9 ± 0.3</td>
<td>96.8 ± 0.2</td>
<td>99.8 ± 0.1</td>
<td>98.5 ± 0.2</td>
</tr>
<tr>
<td>JAN-A</td>
<td>75.2 ± 0.3</td>
<td>96.8 ± 0.2</td>
<td>99.8 ± 0.1</td>
<td>98.5 ± 0.2</td>
</tr>
<tr>
<td>ResNet</td>
<td>66.4 ± 0.2</td>
<td>96.7 ± 0.1</td>
<td>95.8 ± 0.2</td>
<td>94.0 ± 0.2</td>
</tr>
<tr>
<td>JAN</td>
<td>82.9 ± 0.2</td>
<td>97.4 ± 0.1</td>
<td>99.9 ± 0.2</td>
<td>98.5 ± 0.2</td>
</tr>
<tr>
<td>JAN-A</td>
<td>86.0 ± 0.4</td>
<td>97.5 ± 0.1</td>
<td>97.8 ± 0.1</td>
<td>95.5 ± 0.1</td>
</tr>
</tbody>
</table>

WWW: http://ise.thss.tsinghua.edu.cn/~mlong