VGG Reading Group
Learning to Learn: Model Regression Networks for Easy Small Sample Learning

Yu-Xiong Wang and Martial Hebert
Summary

- Addresses learning from small sets of samples, e.g. for few-shot learning or supervised domain adaptation.

- Hypothesizes the existence of a generic transformation $T$ from models learnt on few samples to models learnt of large-samples.

- Proposes a method to regress $T$ using a neural net.

- Shows the method on domain adaptation and one-shot learning.
Main Idea

Notation:

$w^*$ : Classifier from large-sample to $w^*$

$T$ : Transformation from $w^0$ to $w^*$

$T$ is regressed from pairs of $w^0$ and $w^*$
Regression of $T$

- **Feature space**: Pre-trained Alexnet CNN on ILSVRC
- **Classification model**: Linear SVMs
- **Generation of model pairs** $\left\{ (w^0_j, w^*_j) \right\}_{j=1}^J$
  - **Training set**: 700,000 model pairs of 1,000 categories
  - $w^0$: Trained from random small-sample sets $\{(x_i, y_i)\}_{i=1}^M$ with different SVM parameters (data augmentation)
- **Loss function** $L(\Theta)$

$$\sum_{j=1}^J \left\{ \frac{1}{2} \| w^*_j - T(w^0_j, \Theta) \|_2^2 + \lambda \sum_{i=1}^M \left[ 1 - y^j_i \left( T\left( w^0_j, \Theta \right)^T x^j_i \right) \right]_+ \right\}$$
Producing a new model from few samples

Initialization. In this first step, we directly learn the target model $\mathbf{w}^0$ on the small training sample set $\{(\mathbf{x}_i, y_i)\}_{i=1}^K$.

Transformation. Using $\mathbf{w}^0$ as input to our learned model regression network, after forward propagation, we obtain the output model $T(\mathbf{w}^0, \Theta)$. This thus encodes the prior knowledge about $\mathbf{w}$ being preferable.

Refinement. We then introduce $T(\mathbf{w}^0, \Theta)$ as biased regularization into the standard SVM max-margin formulation to retrain the model by minimizing

$$R(\mathbf{w}) = \frac{1}{2} \| \mathbf{w} - T(\mathbf{w}^0, \Theta) \|_2^2 + \eta \sum_{i=1}^K \left[ 1 - y_i (\mathbf{w}' \mathbf{x}_i) \right]_+$$
Experiments: Sanity Check

**Fig. 3.** Performance sanity check of the model regression network by comparing small-sample models $\mathbf{w}^0$, large-sample models $\mathbf{w}^*$ (learned on thousands of examples), and regressed models $T(\mathbf{w}^0)$ on the held-out ILSVRC validation set. X-axis: number of positive training examples. Y-axis: average binary classification accuracy. Our network effectively identifies a generic model transformation
**Experiments: One-shot Domain Adaptation**

**Office dataset: 3 domains**
- Uses AlexNet trained on ILSVRC, and $w$ is a linear SVM.
- Selects the 16 categories common between Office and ILSVRC.
- Testing is done on the Office’s Webcam domain.
- Sample 20 sets of 1 labeled sample in the target domain, and 10 test images. Results are averaged.

<table>
<thead>
<tr>
<th>Prior knowledge</th>
<th>Method</th>
<th>Acc (%)</th>
<th>Prior knowledge</th>
<th>Method</th>
<th>Acc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>SVM (target only)</td>
<td>62.28</td>
<td>Feature</td>
<td>GFK</td>
<td>65.16</td>
</tr>
<tr>
<td>NA</td>
<td>SVM (source only)</td>
<td>53.51</td>
<td></td>
<td>SA</td>
<td>59.30</td>
</tr>
<tr>
<td></td>
<td>SVM (source and target)</td>
<td>56.68</td>
<td></td>
<td>Daumé III</td>
<td>59.21</td>
</tr>
<tr>
<td>Model parameter</td>
<td>PMT</td>
<td>66.30</td>
<td>Joint</td>
<td>MMDT</td>
<td>59.21</td>
</tr>
<tr>
<td></td>
<td>Late fusion (Max)</td>
<td>59.59</td>
<td></td>
<td>Fine-tuning</td>
<td>61.13</td>
</tr>
<tr>
<td></td>
<td>Late fusion (Lin. Int. Avg)</td>
<td>60.64</td>
<td></td>
<td>Regression network (Ours)</td>
<td>68.47</td>
</tr>
</tbody>
</table>
Experiments: Learning Novel Categories

- Features are from AlexNet
- Models are SVMs
- 'Original' is the model trained on the available samples in each dataset
- Results are repeated with other CNN models, and logistic regression instead of SVM.