Abstract

- We propose a novel loss function called Max-Margin Loss that benefits from set-based information by drawing inter-class margins. It maximizes the maximum possible inter-class margin that is calculated by SVMs. It implicitly pushes all the samples towards correct side of the margin with a vector perpendicular to the hyperplane and a strength inversely proportional to the distance to the hyperplane.
- We review existing set-based DML approaches and evaluate them and their combinations together with Max-Margin Loss and Softmax Loss.
- We build a framework where such functions can operate jointly with sample-based ones and investigate the strategies to maintain set information during training.

Related Work

1. Sample-based DML methods (e.g., Softmax, Contrastive, Triplet, LSE, Quadruplet Losses)
2. Set-based DML methods
   (a) Center Loss [4]: Biased centroid estimation and Do not tolerate intra-class variation
   (b) Magnet Loss [3]: Not combinatorial with other functions
3. SVM based Deep Learning [5]: Penalizes only slack variables

Overview of Set-based Learning Framework

Max-Margin Loss

The loss is computed by the following formula:

\[ L_M = \lambda_M \sum_{i=1}^{n} \sum_{j=1}^{m} \max(0, 1 - y_{ij} (w^T x_i + b)) \]

where separating hyperplane for jth class is defined as \( w^T x + b = 0 \) and \( \delta(condition) \) equals to 1 if the condition is satisfied and 0 otherwise.

References


Comparison to State-of-the-art

<table>
<thead>
<tr>
<th>Method</th>
<th>#Training Images</th>
<th>#Ids</th>
<th>Input Size</th>
<th>Network (#Params.)</th>
<th>FL on YTF or LF?</th>
<th>Accuracy on YTF (%)</th>
<th>Accuracy on LF (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepFace [1]</td>
<td>4.4M</td>
<td>4,030</td>
<td>152×152</td>
<td>AlexNet(120M)</td>
<td>No</td>
<td>91.4</td>
<td>97.35</td>
</tr>
<tr>
<td>VGG Face [2]</td>
<td>2.62M</td>
<td>2,622</td>
<td>224×224</td>
<td>VGG11(38M)</td>
<td>Yes</td>
<td>97.3</td>
<td>98.95</td>
</tr>
<tr>
<td>VGG Face [2]</td>
<td>2.62M</td>
<td>2,622</td>
<td>224×224</td>
<td>VGG13(38M)</td>
<td>No</td>
<td>91.6</td>
<td>-</td>
</tr>
<tr>
<td>( L_S + L_M )</td>
<td>0.83M</td>
<td>2,558</td>
<td>96×96</td>
<td>NNS1(26M)</td>
<td>No</td>
<td>92.44</td>
<td>96.03</td>
</tr>
</tbody>
</table>

Table 1: Comparison of Methods to State-of-the-art.