Learning Meta Face Recognition in Unseen Domains
-Supplementary Material-

A. Theoretical Analysis for MFR

Our meta-objective of MFR is summarized as:

$$\arg\min_{\theta} \gamma L_S(\theta) + (1 - \gamma) L_T(\theta - \alpha L'_S(\theta)).$$  (1)

MFR The first order Taylor expansion has the following form:

$$f(x) = f(x_0) + f'(x_0) \times (x - x_0),$$  (2)

where $x_0$ is a value close to $x$. Let $x = \theta - \alpha L'_S(\theta)$ and $x_0 = \theta$, the second term of Eqn. 1 becomes:

$$L_T(\theta - \alpha L'_S(\theta)) = L_T(\theta) + L'_T(\theta) \cdot (-\alpha L'_S(\theta)).$$  (3)

Then the meta objective Eqn. 1 becomes:

$$\arg\min_{\theta} \gamma L_S(\theta) + (1 - \gamma) L_T(\theta) - \alpha (1 - \gamma)(L'_S(\theta)L'_T(\theta)).$$  (4)

It indicates that the model is optimized to: (i) minimize the loss on both meta-train and meta-test domains. (ii) maximize the dot product of $L'_S(\theta)$ and $L'_T(\theta)$. The former is obvious as we want to learn discriminative representations on both domains. For the latter, if we regard $L'_S(\theta)L'_T(\theta)$ as the similarity between two gradient vectors, it can be understood as we want to encourage two gradients on both domains towards a similar direction. Thus this objective can be understood as: optimize the model parameters, such that after updating on the meta-train domains, the model also performs well on the meta-test domain. In contrast, the conventional objective $\arg\min_{\theta} L_S(\theta) + L_T(\theta)$ has no such constraint.

B. Feature Visualization

In Fig. 1, we show t-SNE projections of face representations from testing sets in four racial domains. We compare Base and MFR options in the GFR-V protocol. The visualization shows that our method MFR pushes the representations from different domains close to the center and makes them more "domain-invariant" than Base.

<table>
<thead>
<tr>
<th>Method</th>
<th>Base</th>
<th>Base-Agg</th>
<th>MLDG</th>
<th>MFR (Ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPLFW</td>
<td>89.3</td>
<td>89.49</td>
<td>88.58</td>
<td>90.67</td>
</tr>
<tr>
<td>CALFW</td>
<td>95.19</td>
<td>95.27</td>
<td>94.88</td>
<td>96.67</td>
</tr>
<tr>
<td>CFP-FP</td>
<td>93.68</td>
<td>93.84</td>
<td>92.98</td>
<td>95.3</td>
</tr>
</tbody>
</table>

Table 1: Comparative results on CPLFW, CALFW and CFP-FP.

C. Additional Experiments

We perform additional experiments on CALFW, CPLFW and CFP-FP in Table 1, and all of them show the improvements of our proposed MFGR.