A Compact Embedding for Facial Expression Similarity - Supplementary

Anonymous CVPR submission

Paper ID 2958

Figure 1: Triplet prediction accuracy of individual raters on the FEC test set (blue bars). The red line shows the median accuracy.

1. Triplet prediction accuracy of human raters

To estimate how good humans are at identifying the most similar pair in a triplet, we computed the accuracy values for individual raters based on how often they agree with the maximum-voted label. Figure 1 shows the accuracy values for all the 30 raters who contributed to the FEC test set annotations. The mean and median values are 86.2% and 87.5%.

2. Visualization of FECNet embeddings space

Figure 2 shows a 2D t-SNE [1] visualization of the learned FECNet-16d embedding space using the AffectNet [3] validation set. The color of the bounding box of a face represents the emotion label of that face. Please zoom in to see the face images clearly.

3. Photo album summarization

We created ten photo albums (100-200 images in each album) by downloading images of ten celebrities using Google image search, and generated two visual summaries for each album by clustering the FECNet-16d and AFFNet-CL-P embeddings, respectively. We showed these two summaries to ten human raters and asked them which one is better. For each decision, the winner and looser get a score of 1 and 0, respectively. If it is a tie, then both of them get a score of 0. Table 1 shows the overall scores for the FECNet-16d and AFFNet-CL-P embeddings for all ten albums. As we can see, human prefer the summaries generated by the proposed FECNet-16d embedding for eight out of ten albums. Figures 3 and 4 show the summary results for all ten albums. For most albums, the expression content is more diverse in the summaries generated by the proposed FECNet-16d embedding.

4. Expression retrieval

Figures 5 to 9 show the expression retrieval results (top-10 images from CelebA [2] dataset) of the FECNet-16d and AFFNet-CL-P embeddings for all the 25 expressions in the query set. In each figure, the images on the left are the query expressions. For each query, the top and bottom rows correspond to the FECNet-16d and AFFNet-CL-P embeddings, respectively. We can see that the proposed FECNet-16d embedding produces better matches when compared to the AFFNet-CL-P embedding.

References


Table 1: Scores received by the FECNet-16d and AFFNet-CL-P embeddings in the photo album summarization task.

<table>
<thead>
<tr>
<th>Album</th>
<th>KM</th>
<th>DT</th>
<th>CB</th>
<th>GB</th>
<th>BO</th>
</tr>
</thead>
<tbody>
<tr>
<td>FECNet-16d vs AFFNet-CL-P</td>
<td>10-0</td>
<td>5-1</td>
<td>9-1</td>
<td>9-0</td>
<td>5-2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Album</th>
<th>JL</th>
<th>LS</th>
<th>HC</th>
<th>LJ</th>
<th>JC</th>
</tr>
</thead>
<tbody>
<tr>
<td>FECNet-16d vs AFFNet-CL-P</td>
<td>9-0</td>
<td>1-6</td>
<td>10-0</td>
<td>1-4</td>
<td>7-1</td>
</tr>
</tbody>
</table>
Figure 2: 2D visualization of the learned FECNet-16d embedding space using t-SNE [1]. The color of the bounding box of a face represents the emotion label of that face. Please zoom in to see the face images clearly.
Figure 3: Summaries generated by the FECNet-16d (left) and AFFNet-CL-P (right) embeddings.
Figure 4: Summaries generated by the FECNet-16d (left) and AFFNet-CL-P (right) embeddings.
Figure 5: Expression retrieval results - Images on the left are the query expressions. For each query, the top row corresponds to the FECNet-16d embedding and the bottom row corresponds to the AFFNet-CL-P embedding.
Figure 6: Expression retrieval results - Images on the left are the query expressions. For each query, the top row corresponds to the FECNet-16d embedding and the bottom row corresponds to the AFFNet-CL-P embedding.
Figure 7: Expression retrieval results - Images on the left are the query expressions. For each query, the top row corresponds to the FECNet-16d embedding and the bottom row corresponds to the AFFNet-CL-P embedding.
Figure 8: Expression retrieval results - Images on the left are the query expressions. For each query, the top row corresponds to the FECNet-16d embedding and the bottom row corresponds to the AFFNet-CL-P embedding.
Figure 9: Expression retrieval results - Images on the left are the query expressions. For each query, the top row corresponds to the FECNet-16d embedding and the bottom row corresponds to the AFFNet-CL-P embedding.