Introduction:

- Existing Techniques in Human Facial Image-base Age Estimation:
  - Feature extraction: geometry features, engineered features.
  - Estimators: classification (SVM), regression (SVR), ranking.
  - Deep learning: multi-class CNN, multi-scale CNN, MR-CNN, DEX.

- Ranking-CNN
  - Contains a series of basic CNNs.
  - Initialized with a pre-trained base CNN, fine-tuned with ordinal age labels.
  - The binary outputs are aggregated to make the final age prediction.

Theoretical Analysis:

- A new error bound for ranking

**Theorem** For any observation \((x, y)\), in which \(y > 0\) is the actual label (integer), then the following inequality holds:

\[
|r(x) - y| \leq \max_k e_k(x),
\]

where \(r(x)\) is the estimated rank of age, \(k = 1, \ldots, K - 1\). That is, we can diminish the final ranking error by minimizing the greatest binary error.

- Step I
  - Input \(x\), output \(y\), aggregated rank \(r(x)\).
  - \(E^r\): the number of misclassifications when \(y < r(x)\).
  - \(E^e\): the number of misclassifications when \(y > r(x)\).

- Step II
  - Denote the error produced by each binary ranker as \(e_i, i = 1, \ldots, k\).
  - Sort the sub errors of \(E^r\) in an increasing order and find out the ordinal relation:

\[
e_1 > e_2 > \ldots > e_k \geq 0
\]

- Step III
  - Finally, \(|r(x) - y| \leq e_k\).

Experiments:

- Experiment setup
  - Pre-train with 26,580 image samples from the unfiltered faces dataset.
  - Fine-tune on the age estimation benchmark MORPH dataset.
  - Randomly select 54,362 samples in the age range between 16 and 66.

**Comparison** of MAE among different combinations of features and estimators.

<table>
<thead>
<tr>
<th>CLASSIFICATION</th>
<th>SVM</th>
<th>MULTI-CLASS CNN</th>
<th>Ranking-CNN</th>
<th>OR-CNN</th>
<th>CNN FEATURE</th>
<th>RANKING-CNN FEATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASSIFICATION</td>
<td>SVM</td>
<td>MULTI-CLASS CNN</td>
<td>Ranking-CNN</td>
<td>OR-CNN</td>
<td>CNN FEATURE</td>
<td>RANKING-CNN FEATURE</td>
</tr>
<tr>
<td>MODEL</td>
<td>4.99</td>
<td>3.15</td>
<td>3.05</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RANKING MODEL</td>
<td>3.65</td>
<td>4.43</td>
<td>3.65</td>
<td>2.96</td>
<td>3.65</td>
<td>-</td>
</tr>
</tbody>
</table>

Comparison on Cumulative Score and binary accuracy.

**Explanatory example:** When \(y = 3\), the binary outputs are supposed to be 11000. If we get 00101:
- \(e_1 = 3 + 3 + 1 + 1\), \(e_2 = 5 + 3 + 1 + 3\), \(|r(x)| = 0\), \(E^e = 2\), \(e_k = e_3 = 3\).

Acknowledgement

- NSF: CNS 1637312
- Ford URP:2015-9186R

https://github.com/RankingCNN
http://www.cs.wayne.edu/~mdong