Ordinal Scale Evaluation of Smiling Intensity using Comparison-Based Network

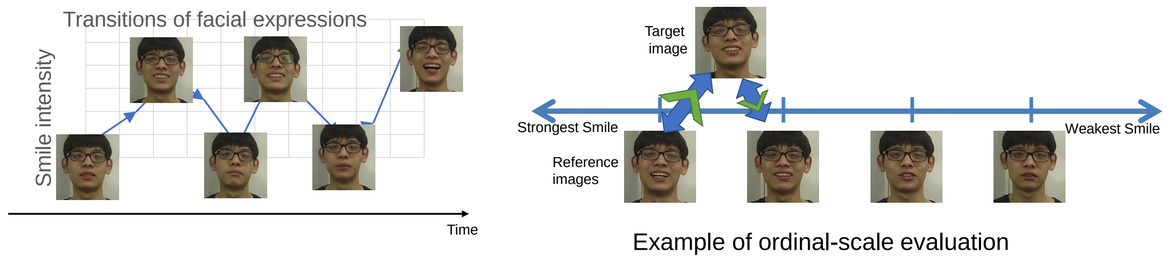
Kei SHIMONISHI, Kazuaki KONDO, Hirotada UEDA, Yuichi NAKAMURA Kyoto University
Contact: shimonishi@i.kyoto-u.ac.jp

Introduction

Objective: Monitoring individual's facial expressions to trace QOL

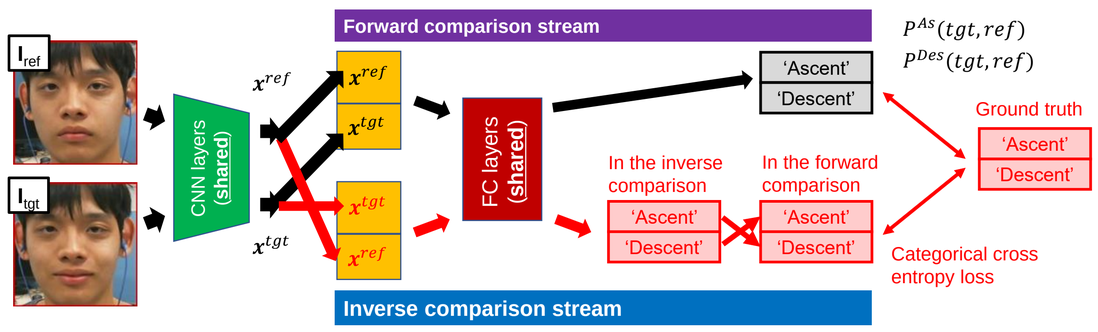
Difficulty: Traditional facial expression recognitions do not give appropriate score for intermediate expressions due to the training method.

Idea: With several reference images and facial expression comparison technique, new expression can be evaluated in the evaluation space for each individual. → **Ordinal-scale** evaluation



Example of ordinal-scale evaluation

Comparison-Based Network

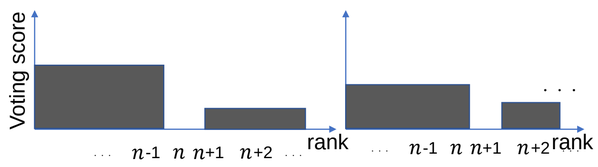


Comparison-based facial expression recognition network [Kondo et al., 2020]
- Input: Two face images
- Output: Which of two face images shows more expressions

Voting-Based Evaluation

Problem in the estimation

Searching r s. t. $sn(I_r) > sn(I) > sn(I_{r+1})$, where $sn(I)$ shows the degree of smiling of face image I, provides unstable estimation because of the ambiguity and inconsistency in the ranking.

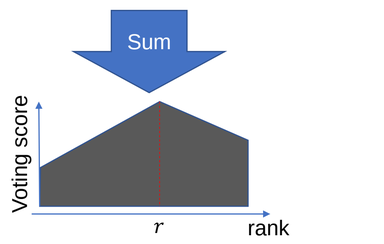


Approach

Using the result of each comparison of reference image I_n and target image I_{new}, voting to possible ranks.
The most likely rank should have maximum # of votes.

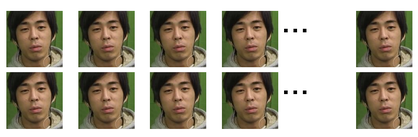
Method

If $sn(I_n) > sn(I_{new})$ is estimated, the smiling intensity rank of I_{new} is lower than n.
In practice, add a likelihood values of '>' and '<' to the ranks lower than and higher than n, respectively.



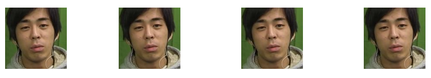
$$r = argmax_r \left\{ \sum_{n=1}^{r-1} P^{Des}(new, n) + \sum_{n=r+1}^{N} P^{As}(new, n) \right\}$$

Ordinal Scale Evaluation of Smiling Intensity



Candidates of reference images

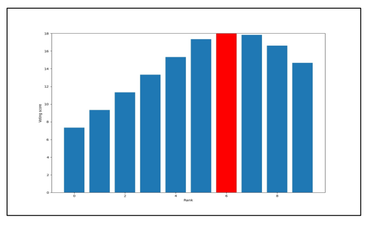
1. Reference image selection



Reference images consisting of the coarse ranking

Target image

2. Comparing and voting



3. Rank of target image

Issue

Since proposed Scheme is an ordinal-scale evaluation, estimation results are strongly affected by the reference images.

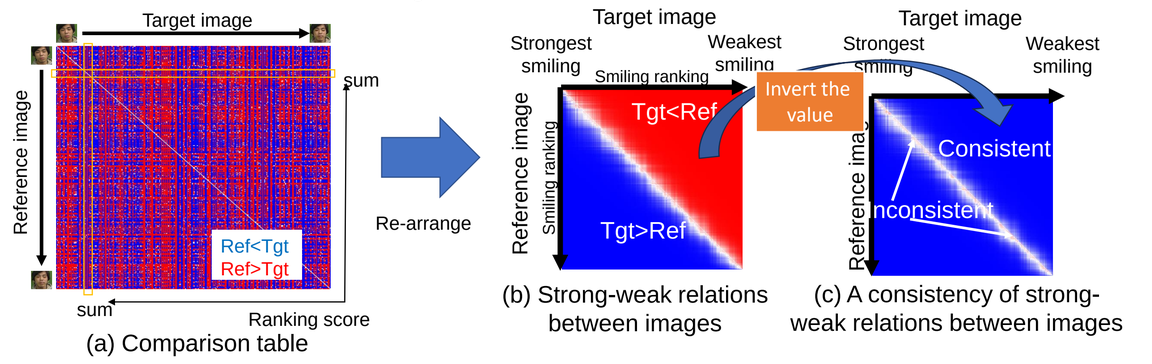
Idea

Reference images are selected from a large number of candidates taking into account consistency of strong-weak relationship.

Reference-Image Selection

Baseline Ranking Construction

1) Apply pairwise-comparison to all images in the dataset and obtain strong-weak relation table.
2) Re-arrange images by sorting the sum of the values in the table to maximize consistency in the ranking.
→ Considered as a baseline ranking



(a) Comparison table
(b) Strong-weak relations between images
(c) A consistency of strong-weak relations between images

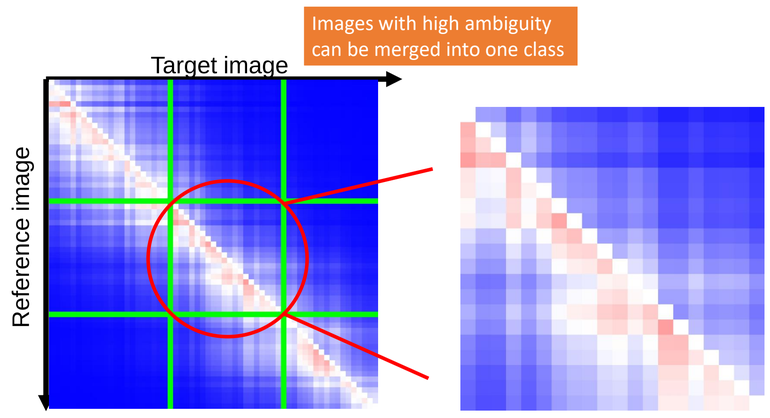
Reference-Image Selection

Idea: Ambiguous images in baseline ranking are merged into one class and considered as "Don't care".

- A consistency square whose values are small should be merged → minimize values in consistency square
- A neighboring images in baseline ranking should not be selected to pick different reference images → maximize the sum of area of consistency square
→ Selecting images to divide baseline ranking by calculating :

$$argmax \frac{\sum(1 - consisy\ values)}{\sum area}$$

Dynamic programing is applied to reduce calculation costs.



(a) A consistency of strong-weak relations in the baseline ranking images
(b) A consistency square of neighboring images

Evaluation

Purpose

To evaluate how well
- the proposed network can compare face images
- the reference images can be selected
- the face images can be evaluated

Scheme

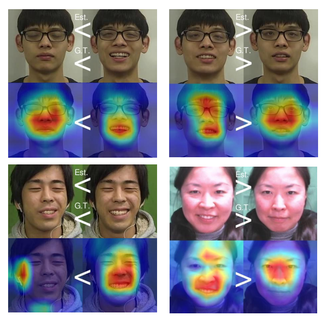
1) Network was trained and evaluated by cross validation
2) Baseline ranking was constructed by the training data
3) Reference images were selected from baseline ranking
4) Human annotators evaluated the reference images
5) Face images obtained by the movie were evaluated

Results

Prediction accuracy

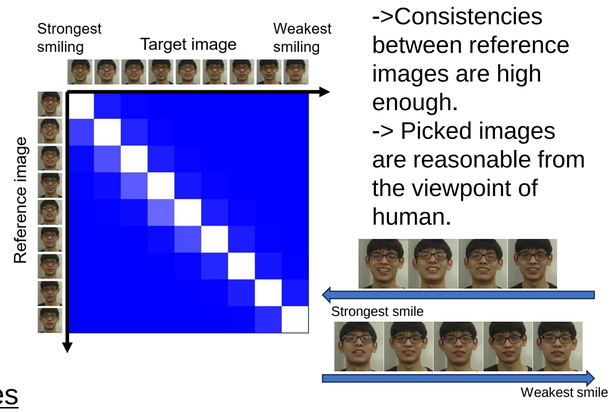
Evaluated by 5-fold cross validation
For three dataset:
- 99.5% (215/216)
- 100% (174/174)
- 98.3% (121/123)
-> High enough

GradCAM
- Region-of-focus contributed to the estimation results.
-> Around mouth and eye regions are correctly focused.



Selected reference images

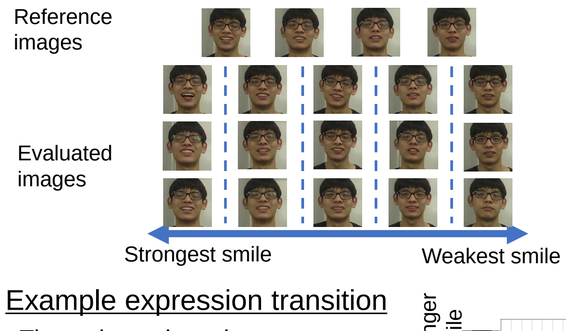
9 images picked from 216 x 2 = 432 images in the baseline ranking are shown below.



->Consistencies between reference images are high enough.
-> Picked images are reasonable from the viewpoint of human.

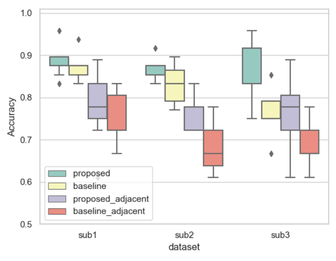
Example evaluation results by ordinal scales

- Since adjacent classes are still similar and it is difficult to evaluate, 4 reference images skip one rank at a time are picked, and example images are listed.
-> Seems to be correctly evaluated.



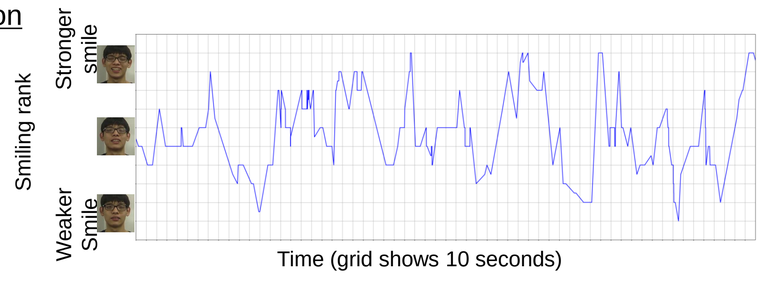
Human evaluation for selected reference images

- 7 human annotators evaluated which of two face images shows more smile in reference images.
- Baseline: equally picked images in each 10% of ranking.
-> Higher accuracy means that the ambiguity of reference images are small.



Example expression transition

- The estimated results are smoothed by median filter.
-> Smiles of slightly stronger intensity occurred several times in succession in the first half of this period.
-> Strongest smiles occurred with a short interval in the second half of this period.



Conclusion

We propose an approach to evaluate the degree of smiling of individuals by ordinal scales based on multiple comparisons. To evaluate by ordinal scales, we also propose an algorithm to select reference images. Experimental results show the capability of our approach.

Future work

To enhance a reliability of our framework, we would like to map our ordinal-scale based framework to some physical index such as facial muscles measured by myoelectricity.