

Supplementary Document for EyeOpener: Editing Eyes in the Wild

1 Document Description

In this supplementary document, we present additional results. Please refer to the corresponding sections in the main paper for discussions and analysis.

In Fig.1, we show a few examples with collected human evaluation scores from Amazon Mechanical Turk. Please find details of evaluation collection in Section 4 of main paper.

In Fig.2, we show additional comparisons between our BFGP image compositing method with other methods. Our method generate more natural looking results under different illumination conditions. Please find details of our algorithm in Section 3 of main paper.

From Fig.3 to Fig.13, we show additional examples of predicting compatible *references* for a given *target* in a personal album. The green number on top of result images indicates the score of realism predicted by our regression model as described in the main paper. The orange number below is rating collected from Amazon Mechanical Turk. The higher the score, the more realistic the result is. 4: definitely realistic ; 3: probably realistic; 2: probably unrealistic and 1: definitely unrealistic. In our experiments, training data dose not include any images of the testing subjects. Please check Section 4 for details of our Amazon Mechanical Turk experimentation and Section 5 for the learning of predictive model.

In Fig.14, we show a few examples on large size input pairs. Our system is robust to the input image size, no resolution-related artifacts observed in our experiments.

Fig.15 shows few examples of cross-identity results on color image input.

We conducted a visual inspection of 2,000 randomly sampled results in our dataset. We assigned to each result one or more of the following 5 tags: (a) realistic results; (b) with pose warping artifacts; (c) with unmatched illumination artifacts; (d) with unrealistic expression and (e) look unrealistic due to other reasons. 48.6% results are tagged (a), 25.1% are tagged (b), 25.8% are tagged (c), 15.9% are tagged (d) and 4.3% are tagged (e). In Fig. 16, we show a few examples with these tags.

In addition to SVR, we also trained Random Forests regression[Breiman 2001] on our collected dataset. The Random Forests model with combination features ψ_2^C has 0.67 AP(similar but slightly worse than SVR, which had 0.70 AP) and RMSE of 1.10 (the RMSE for SVR is 1.07). In Fig. 17, we show comparison of the precision-recall curve of the random forests model and SVR model with ψ_2^C feature.

References

- BREIMAN, L. 2001. Random forests. *Machine learning* 45, 1, 5–32.
- SUNKAVALLI, K., JOHNSON, M. K., MATUSIK, W., AND PFISTER, H. 2010. Multi-scale image harmonization. In *ACM Transactions on Graphics (TOG)*, vol. 29, ACM, 125.
- TAO, M. W., JOHNSON, M. K., AND PARIS, S. 2013. Error-tolerant image compositing. *International journal of computer vision* 103, 2, 178–189.

Score 4



Score 3



Score 2



Score 1



Figure 1: Human evaluation scores of our eye opening results. Score definitions: 4 - definitely realistic; 3 - probably realistic; 2 - probably unrealistic; 1 - definitely unrealistic.

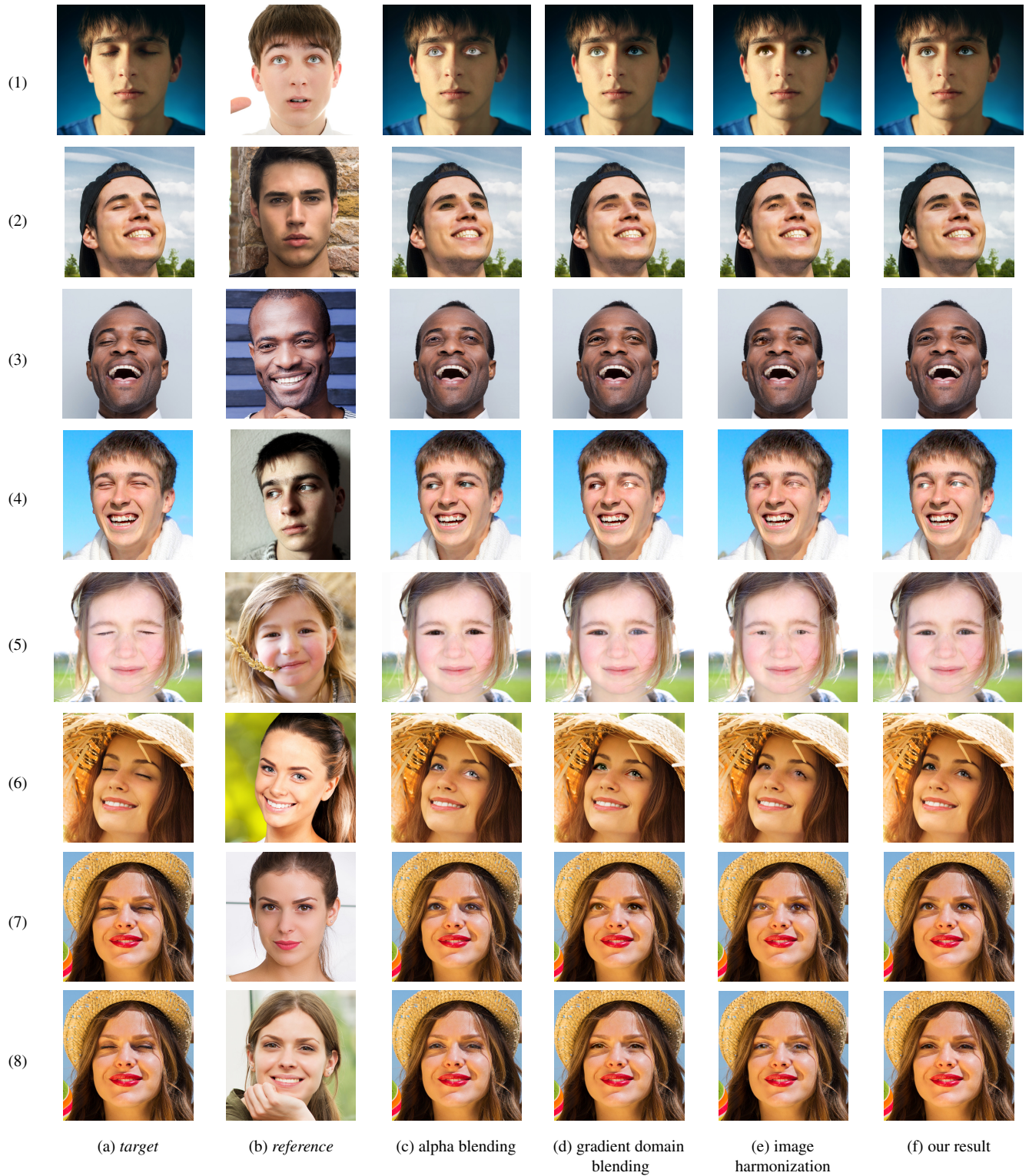


Figure 2: A comparison between different image compositing methods. (a) target image; (b) reference image; (c) result using alpha blending; (d) result using Error Tolerant Gradient-domain Blending[Tao et al. 2013]; (e) result by Image Harmonization[Sunkavalli et al. 2010]; (f) our result. Our method is more robust to differences in illumination conditions.

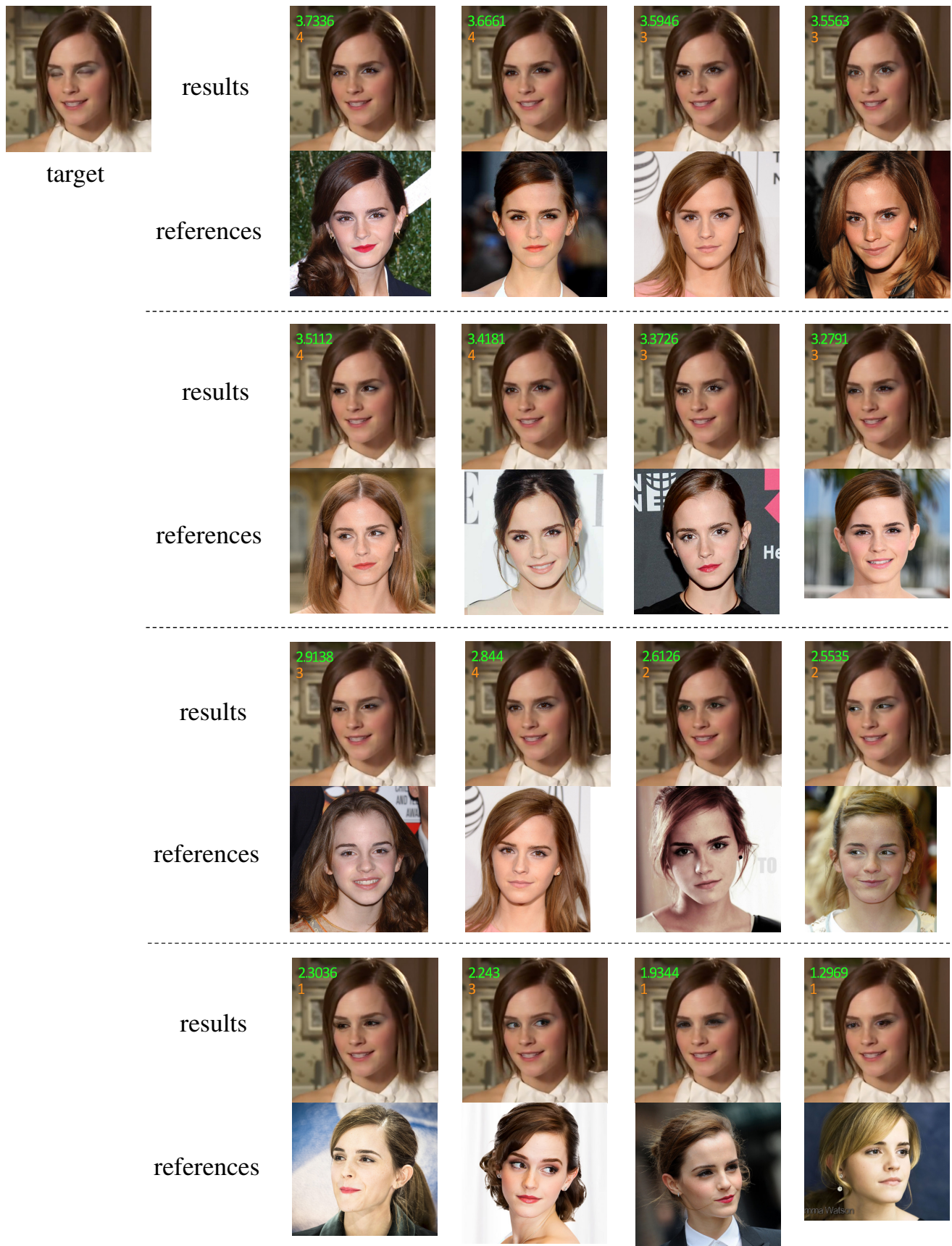


Figure 3: Predict compatible references for a given target in a personal album. The green number on top of result image is predicted score. The orange number below is human evaluation. The higher the score, the more realistic the result is. 4: definitely realistic ; 3: probably realistic; 2: probably unrealistic and 1: definitely unrealistic.

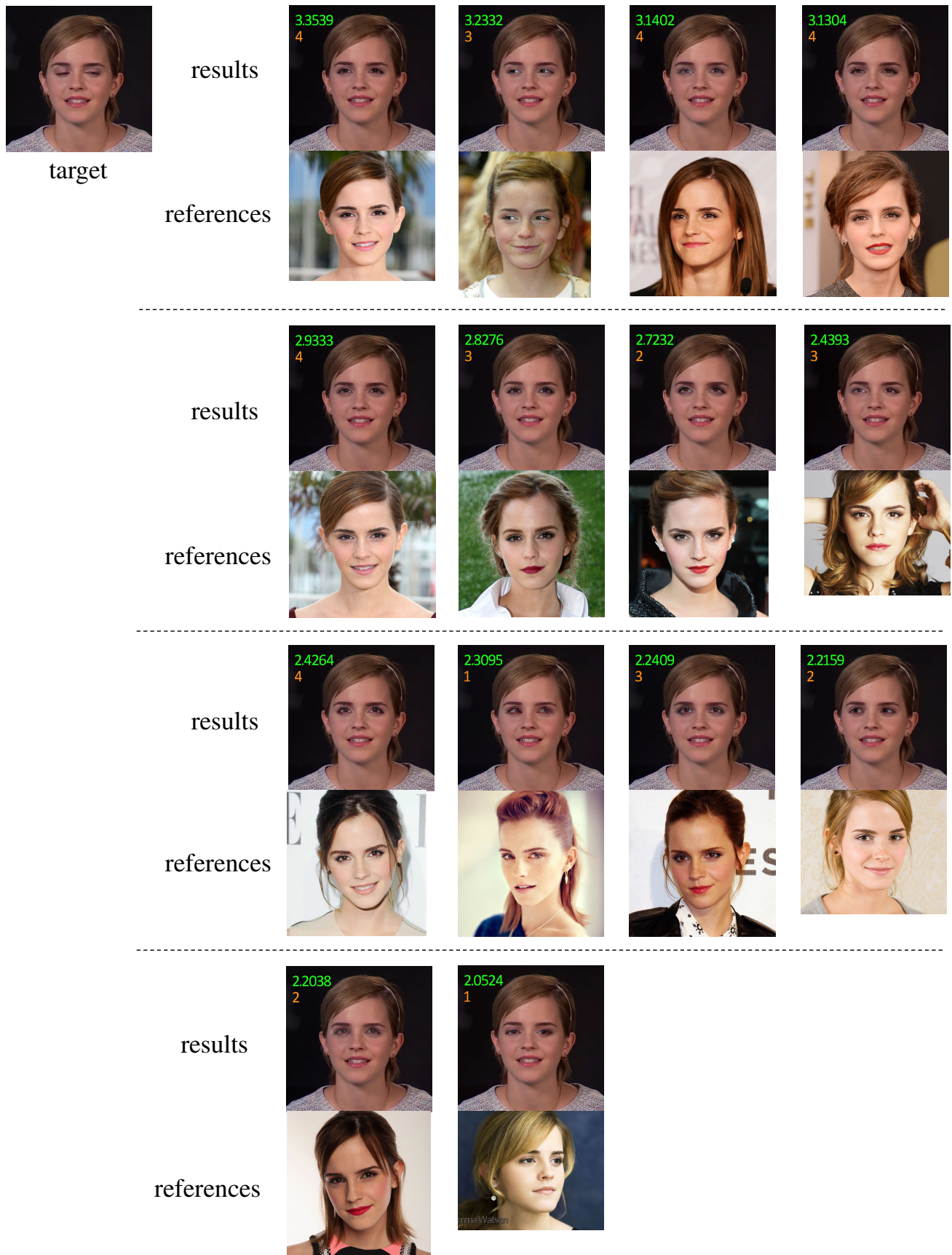


Figure 4: Predict compatible references for a given target in a personal album. The green number on top of result image is predicted score. The orange number below is human evaluation. The higher the score, the more realistic the result is. 4: definitely realistic ; 3: probably realistic; 2: probably unrealistic and 1: definitely unrealistic.

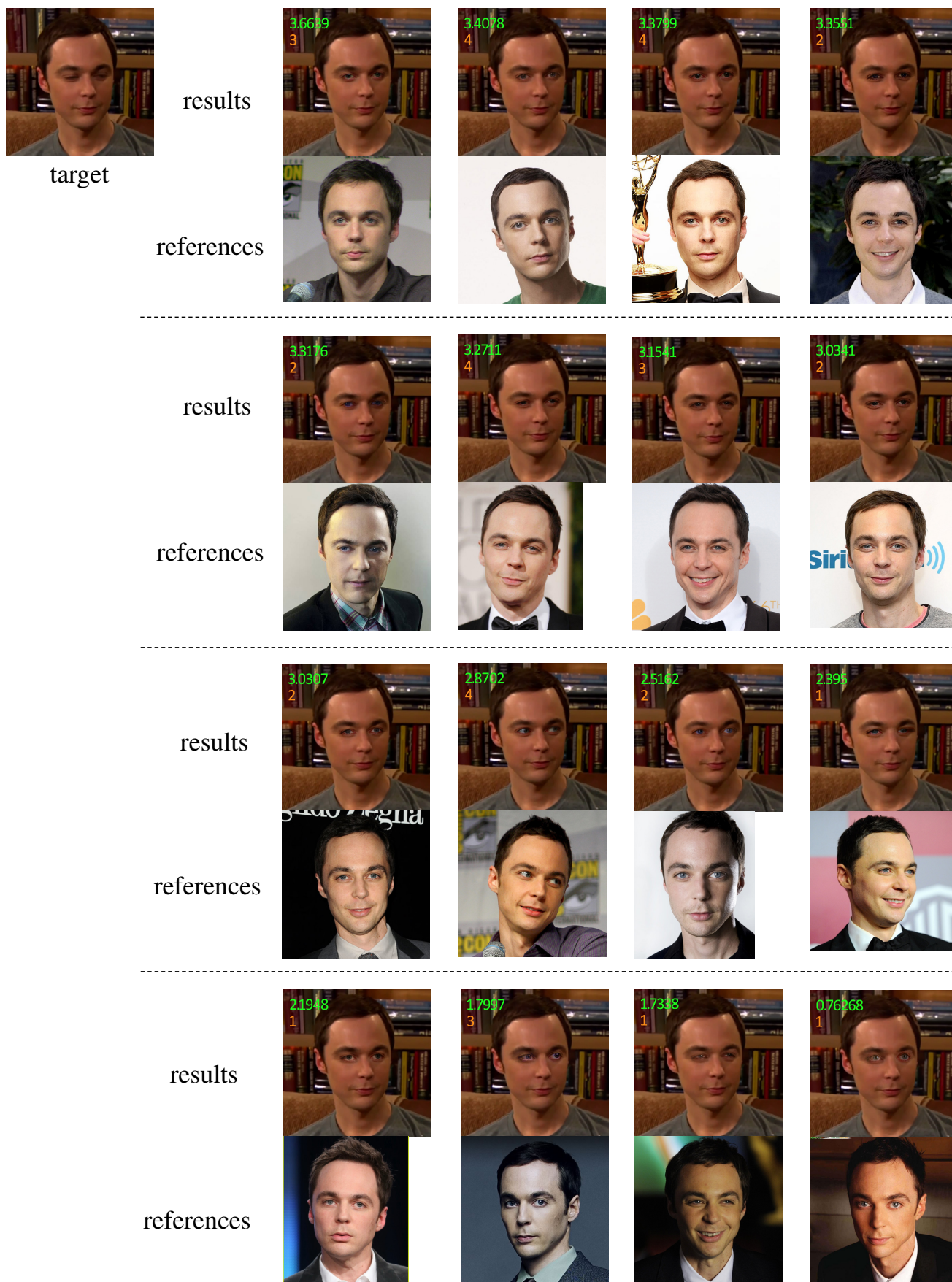


Figure 5: Predict compatible references for a given target in a personal album. The green number on top of result image is predicted score. The orange number below is human evaluation. The higher the score, the more realistic the result is. 4: definitely realistic ; 3: probably realistic; 2: probably unrealistic and 1: definitely unrealistic.

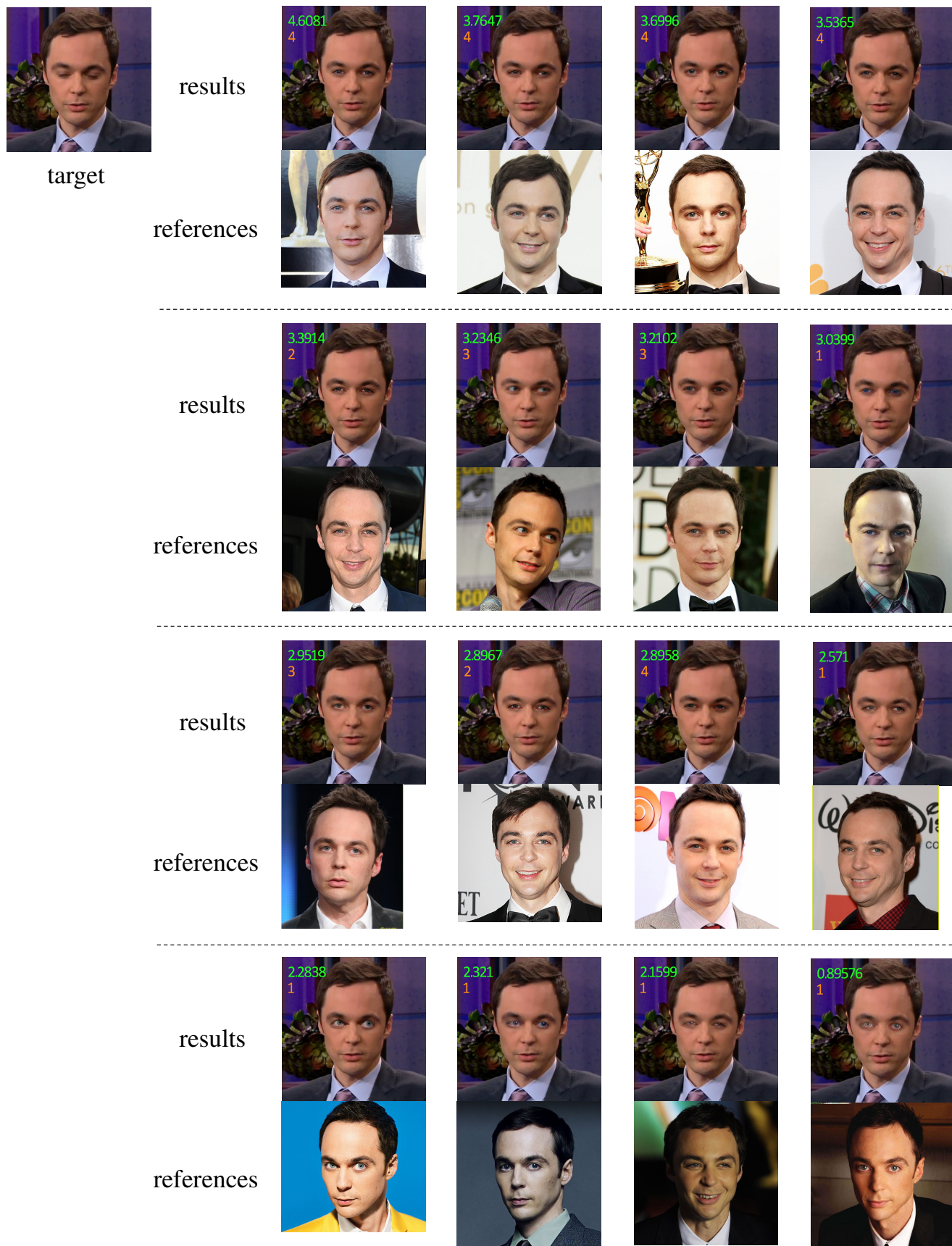


Figure 6: Predict compatible references for a given target in a personal album. The green number on top of result image is predicted score. The orange number below is human evaluation. The higher the score, the more realistic the result is. 4: definitely realistic ; 3: probably realistic; 2: probably unrealistic and 1: definitely unrealistic.

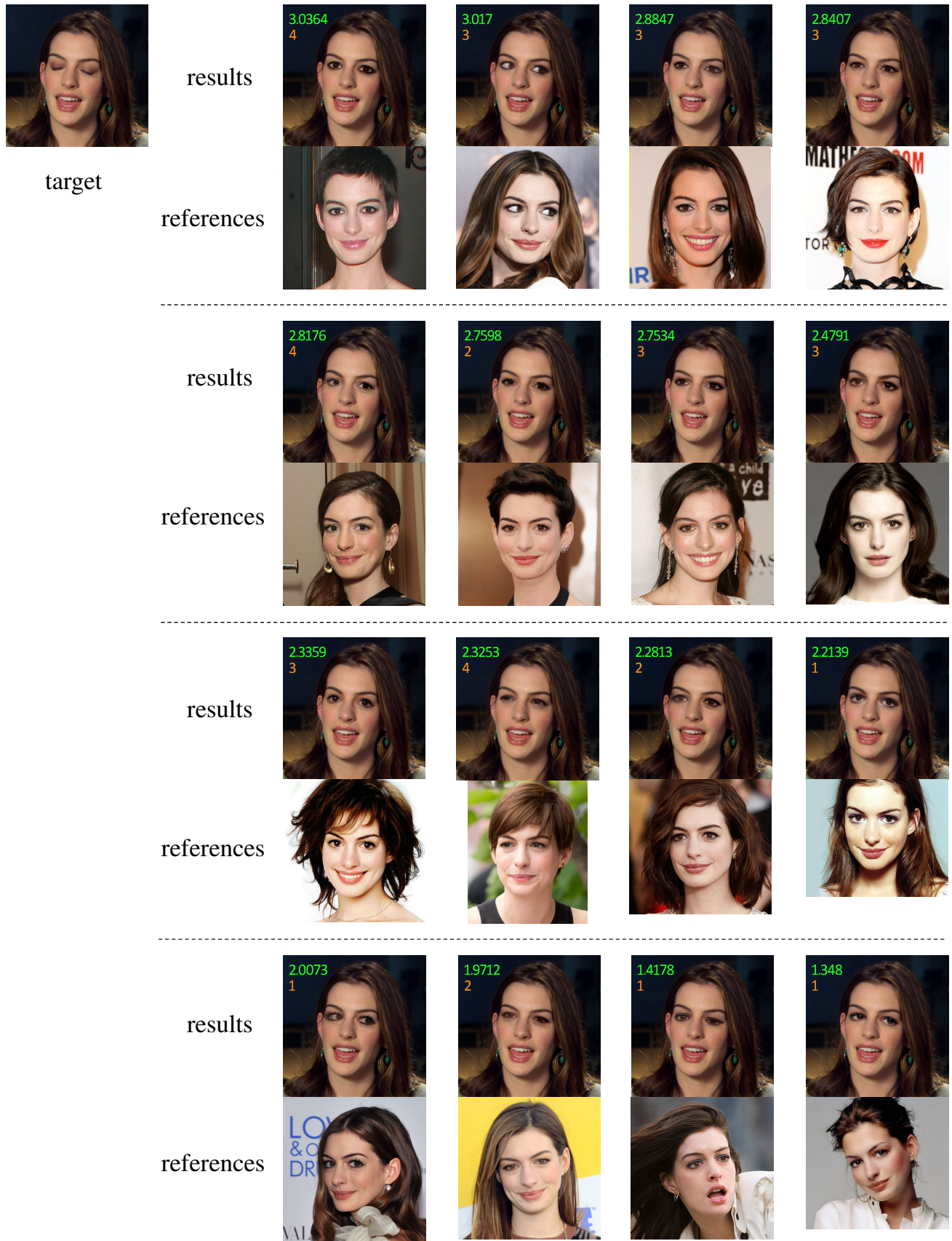


Figure 7: Predict compatible references for a given target in a personal album. The green number on top of result image is predicted score. The orange number below is human evaluation. The higher the score, the more realistic the result is. 4: definitely realistic; 3: probably realistic; 2: probably unrealistic and 1: definitely unrealistic.

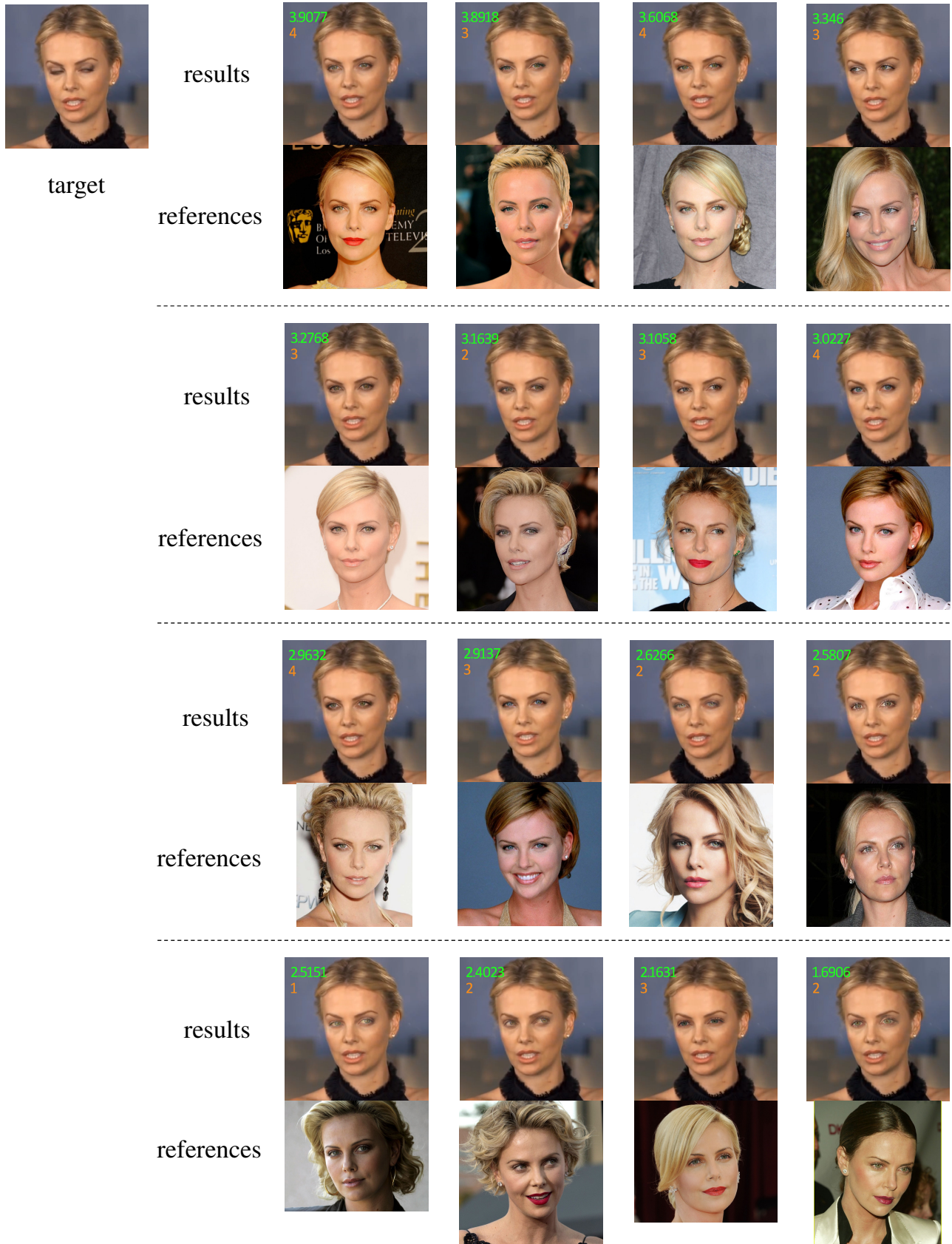


Figure 8: Predict compatible references for a given target in a personal album. The green number on top of result image is predicted score. The orange number below is human evaluation. The higher the score, the more realistic the result is. 4: definitely realistic ; 3: probably realistic; 2: probably unrealistic and 1: definitely unrealistic.



Figure 9: Predict compatible references for a given target in a personal album. The green number on top of result image is predicted score. The orange number below is human evaluation. The higher the score, the more realistic the result is. 4: definitely realistic ; 3: probably realistic; 2: probably unrealistic and 1: definitely unrealistic.



Figure 10: Predict compatible references for a given target in a personal album. The green number on top of result image is predicted score. The orange number below is human evaluation. The higher the score, the more realistic the result is. 4: definitely realistic ; 3: probably realistic; 2: probably unrealistic and 1: definitely unrealistic.

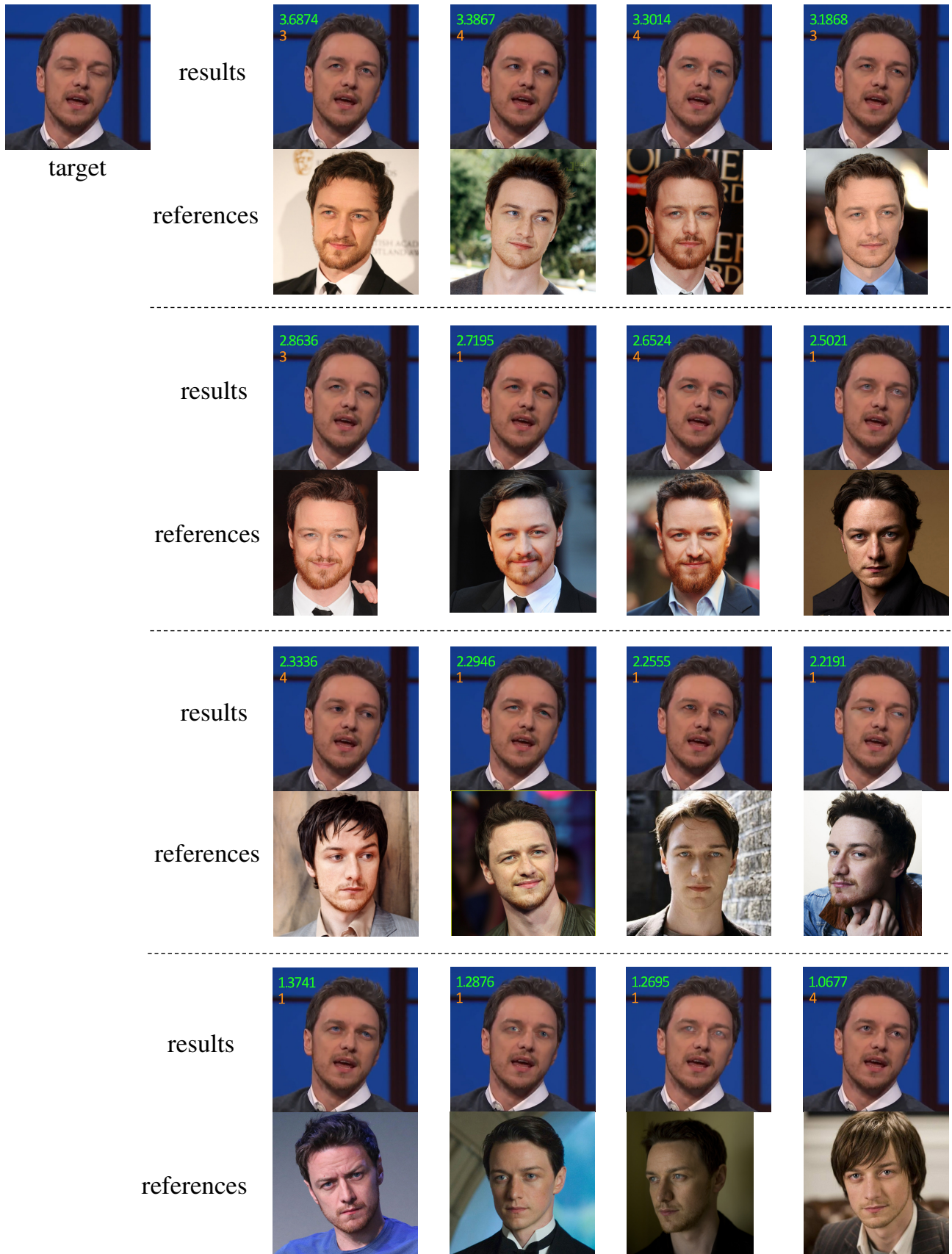


Figure 11: Predict compatible references for a given target in a personal album. The green number on top of result image is predicted score. The orange number below is human evaluation. The higher the score, the more realistic the result is. 4: definitely realistic ; 3: probably realistic; 2: probably unrealistic and 1: definitely unrealistic.

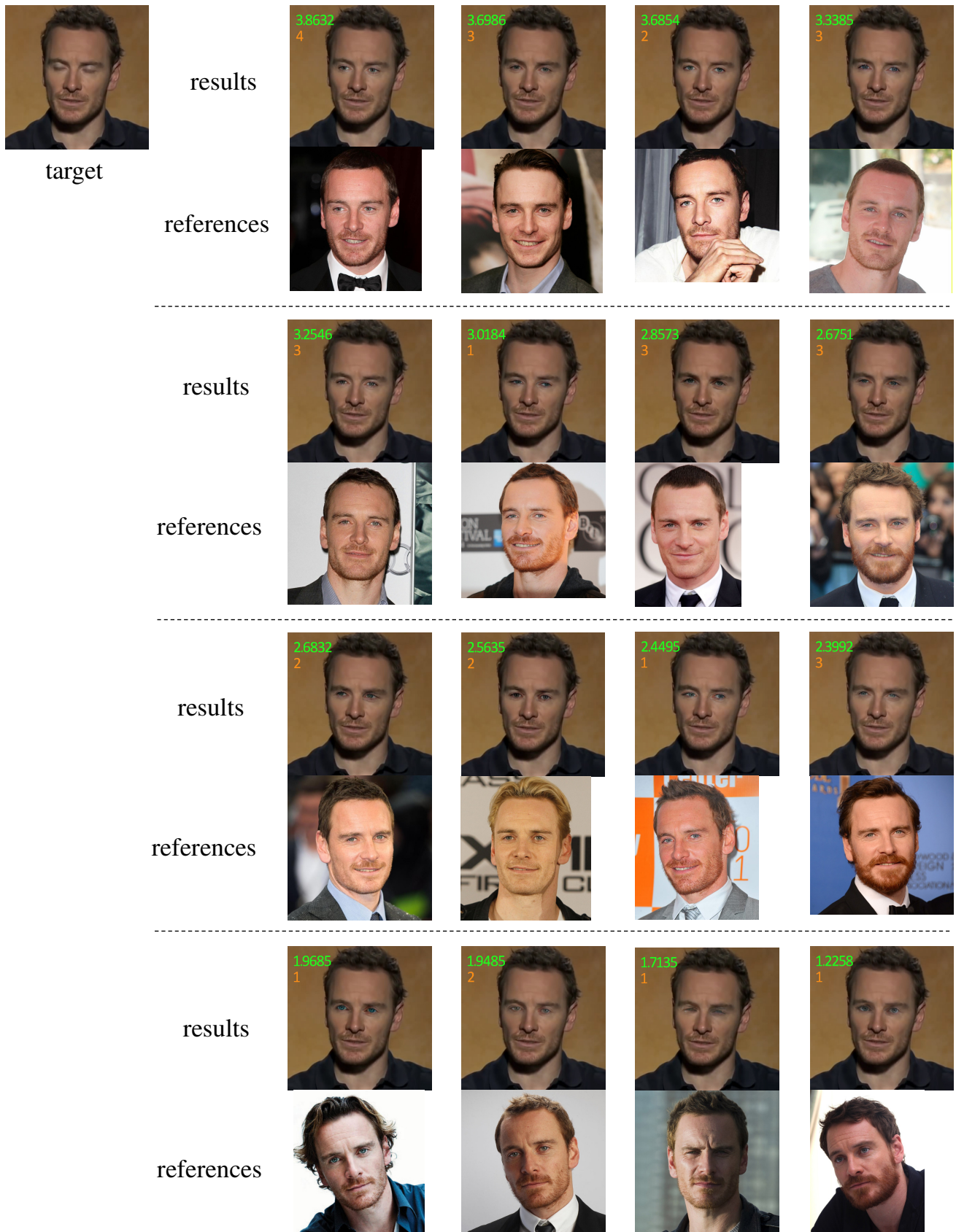
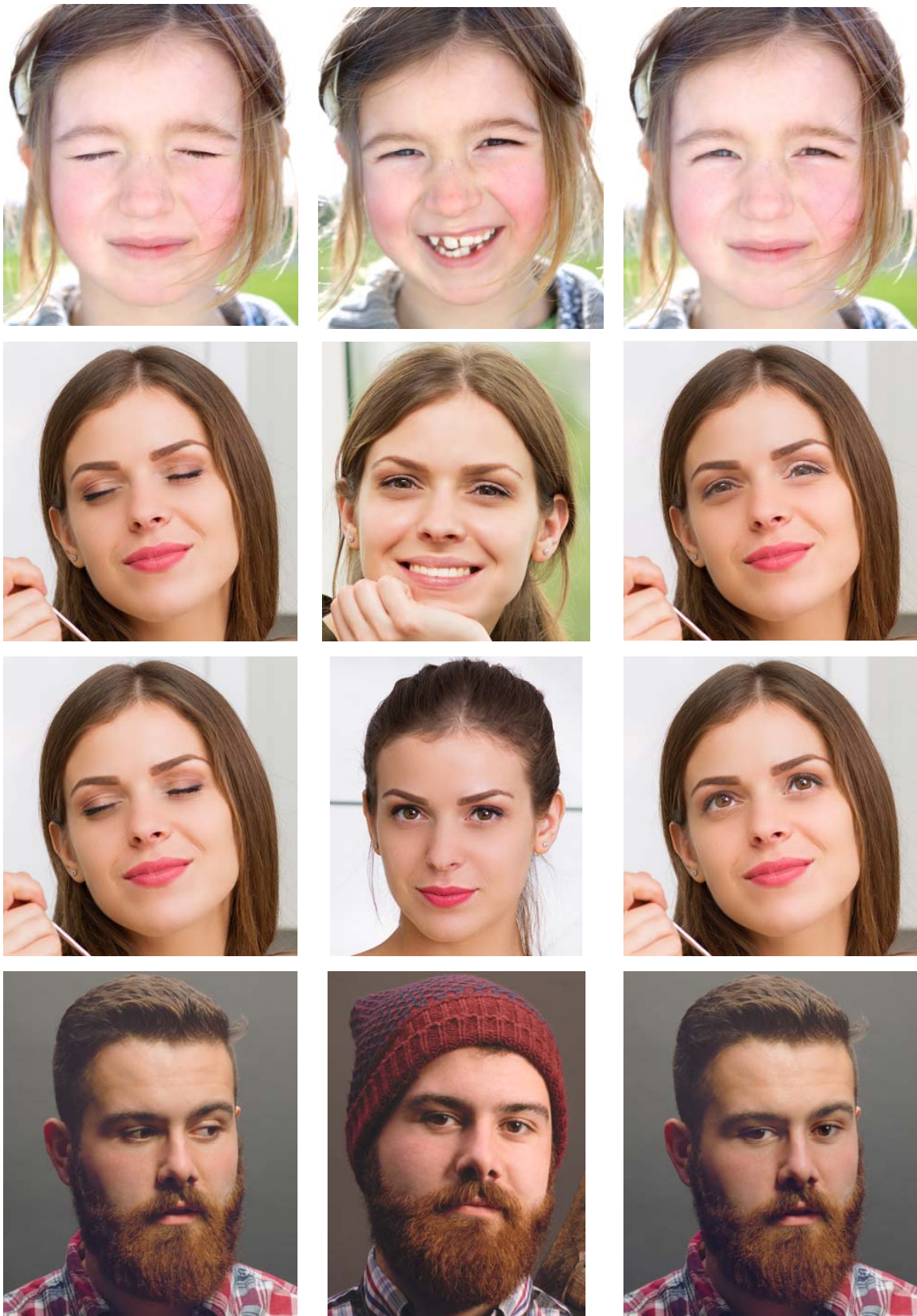


Figure 12: Predict compatible references for a given target in a personal album. The green number on top of result image is predicted score. The orange number below is human evaluation. The higher the score, the more realistic the result is. 4: definitely realistic ; 3: probably realistic; 2: probably unrealistic and 1: definitely unrealistic.



Figure 13: Predict compatible references for a given target in a personal album. The green number on top of result image is predicted score. The orange number below is human evaluation. The higher the score, the more realistic the result is. 4: definitely realistic; 3: probably realistic; 2: probably unrealistic and 1: definitely unrealistic.



(a) targets

(b) references

(c) results

Figure 14: Eye-editing on high resolution input images.

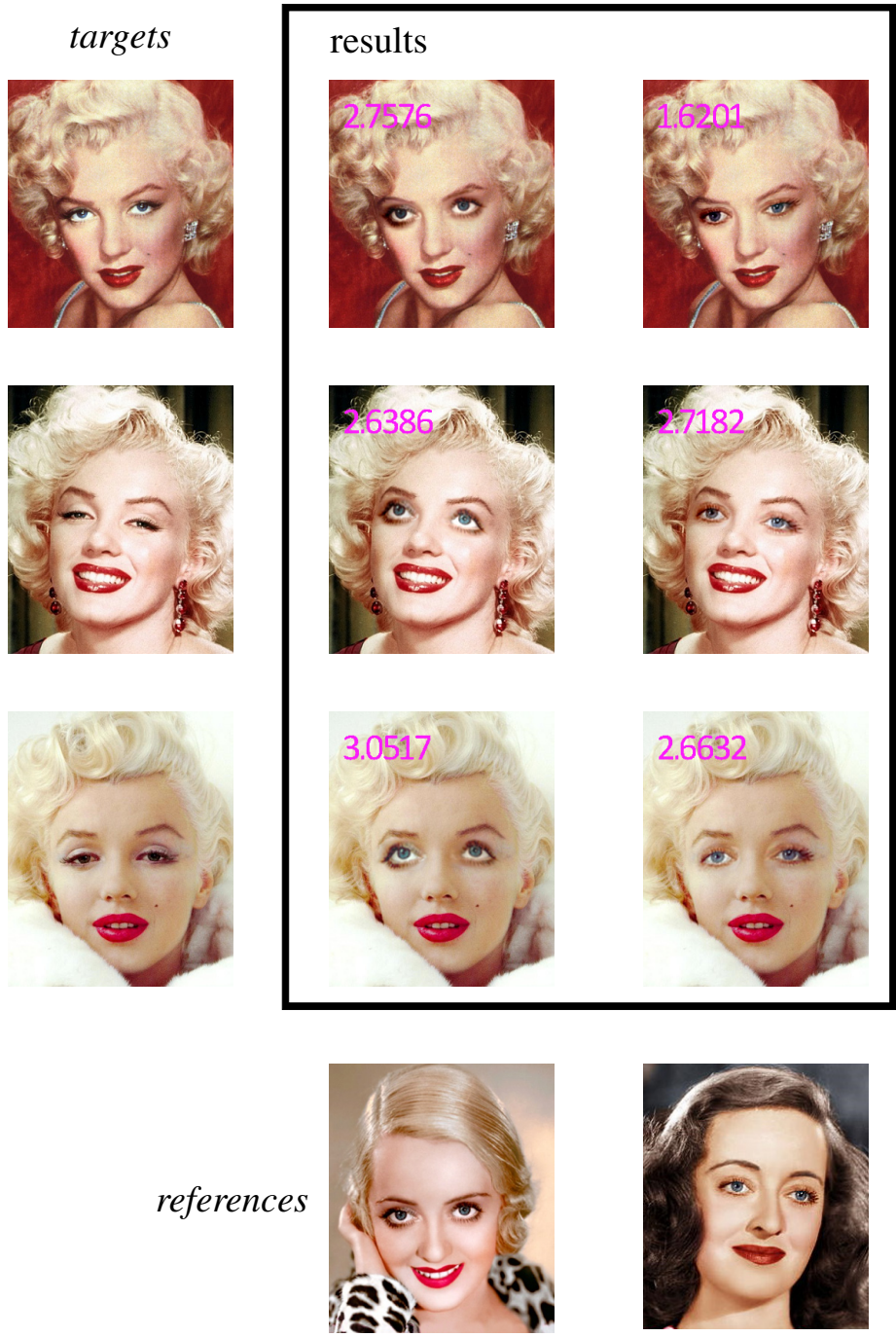
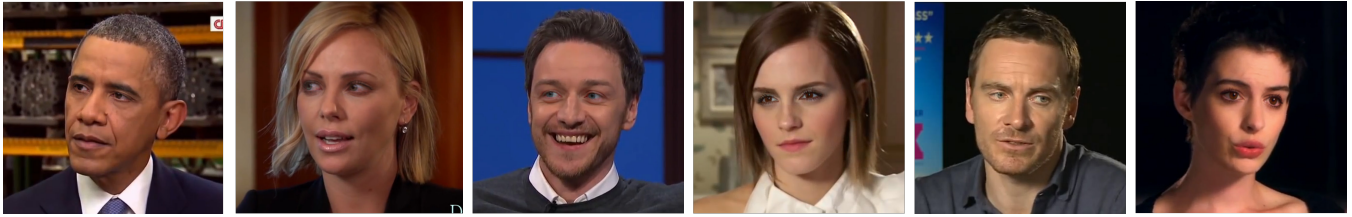


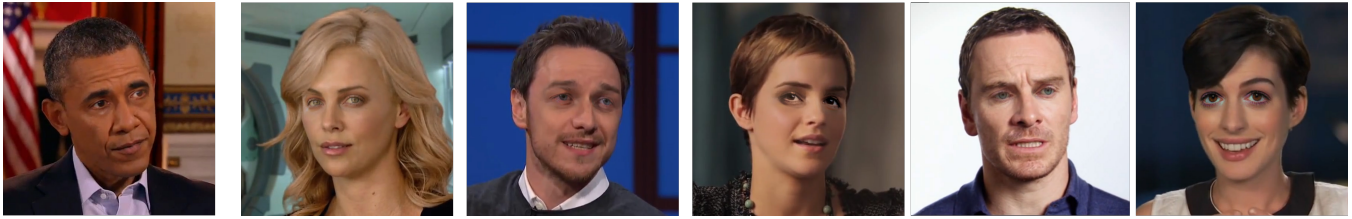
Figure 15: Cross-identity results on color images. Left column: target images. Bottom row: reference images. Each result is generated by our system with the target at the row and the reference at the column.



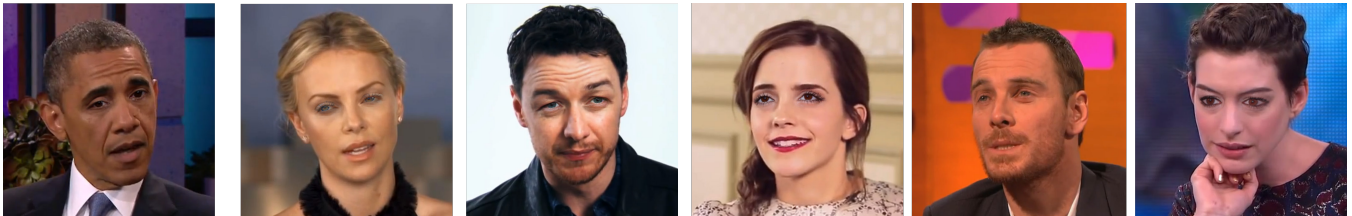
(a) Examples of results considered realistic.



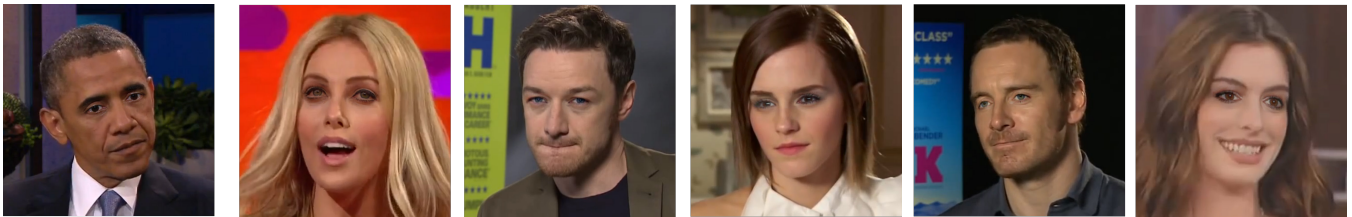
(b) Examples of results contain pose warping artifacts.



(c) Examples of results possess unmatched illumination artifacts.



(d) Examples of results exhibit unrealistic eye expression.



(e) Examples of results considered unrealistic caused by other reasons.

Figure 16: Examples of visual inspection results. We conducted visual inspection on 2,000 randomly sampled results in our dataset. We assigned to each result one or more of the following 5 tags: (a) realistic results; (b) with pose warping artifacts; (c) with unmatched illumination artifacts; (d) with unrealistic expression and (e) look unrealistic due to other reasons. 48.6% results are tagged (a), 25.1% are tagged (b), 25.8% are tagged (c), 15.9% are tagged (d) and 4.3% are tagged (e).

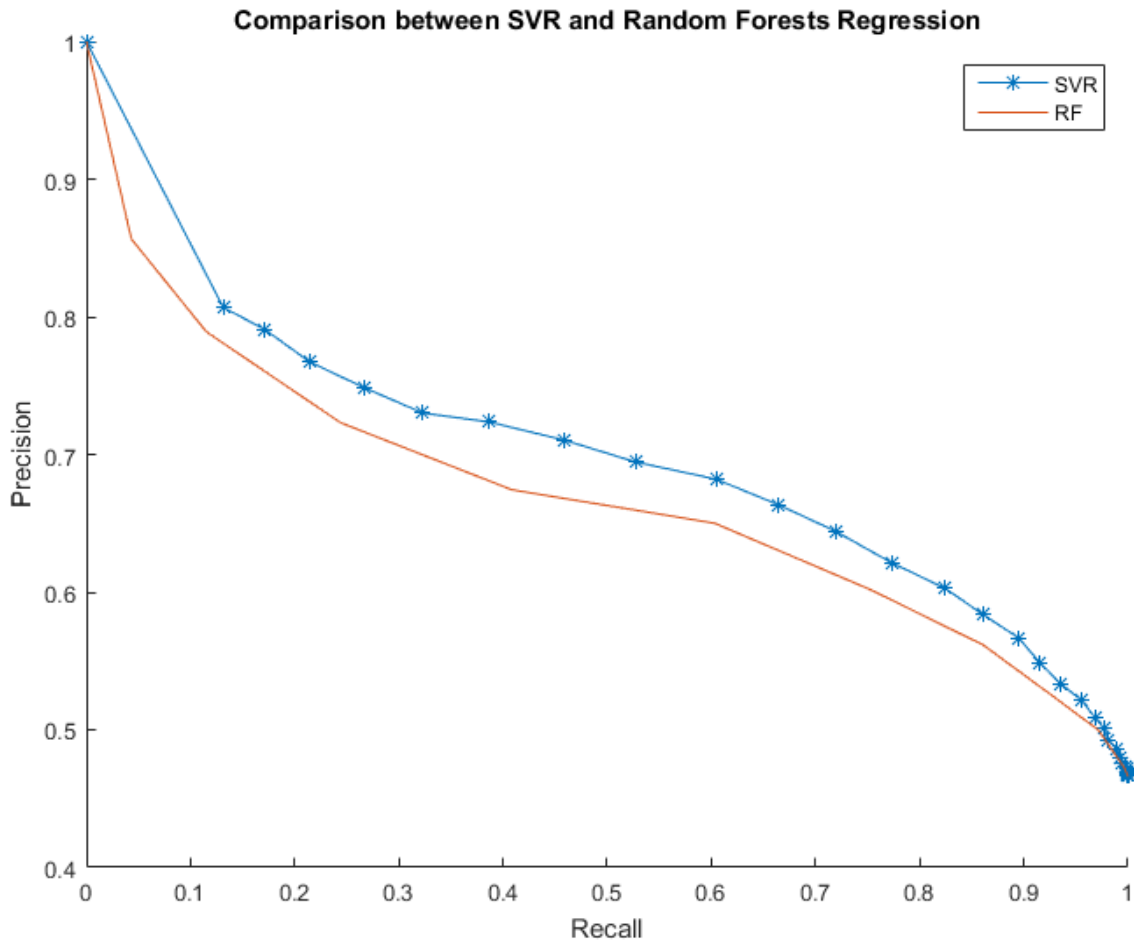


Figure 17: Precision-recall curve of Random Forests regression and Support Vector regression on best features(both are ψ_2^C) respectively.