Asymmetric Image Retrieval with Cross Model Compatible Ensembles
—Supplementary—

1. Experiment details

In this section we provide additional details on the experiments described in Section 4 of the main paper.

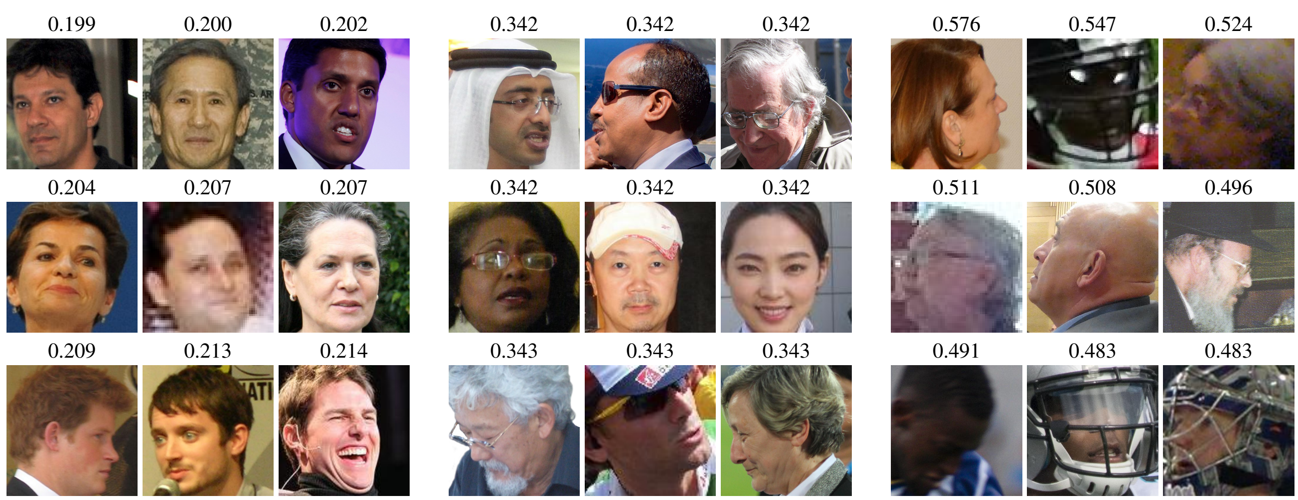
1.1. Implementation details

To train the models, either for gallery or query usage, we used three well-known architectures: ResNet18, ResNet100 [2] and MobileFaceNet [3]. The networks were trained with the *CosFace* [5] loss function and a 512-dimensional embedding space.

To train the transformation models, either for the M2M approach or the Unified approach, we followed Wang *et al*. [4] and implemented the transformation models using four RBT modules [4]. For both approaches, we used the three loss functions described in Section 3 of Wang *et al*. [4]. All models were trained using the VggFace2 dataset [1].

1.2. IJB-C Evaluation protocols

As described in the main text, we evaluate the visual search methods described in the paper using the IJB-C benchmark, which contains 130k images from 3,531 identities. We used the common testing protocols for face recognition, 1:1 verification and 1:N search (“test 1” and “test 4” IJB-C protocols). Since the 1:N protocol (test 4) includes two gallery sets, we evaluated the TAR@FAR=1% results for each gallery individually, and reported the average of these results. In the uncertainty experiments (Section 4.5 in the main paper) only a single gallery set (gallery 1) was used.



(a) Low level variance. (b) Medium level variance. (c) High level variance.

Figure 1. Examples of gallery images with (a) low, (b) medium and (c) high variance values. The numbers above the images shows the corresponding variance value.

2. Images with different uncertainty levels

In Fig. 1, we provide a few examples for gallery images with different uncertainty levels. Gallery images with low uncertainty levels generally contain faces that are clearly visible, as opposed to images with high uncertainty that often contain occluded or extremely rotated faces. Additionally, it can be observed that for a medium level of uncertainty a large proportion of faces wear glasses.

References

[1] Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 67–74, 2018. 1

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[3] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015. 1

[4] Chien-Yi Wang, Ya-Liang Chang, Shang-Ta Yang, Dong Chen, and Shang-Hong Lai. Unified representation learning for cross model compatibility. *arXiv preprint arXiv:2008.04821*, 2020. 1

[5] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5265–5274, 2018. 1