Towards Visual Saliency Explanations of Face Verification Supplementary Material

Yuhang Lu, Zewei Xu, and Touradj Ebrahimi
EPFL, Lausanne, Switzerland
firstname.lastname@epfl.ch

This is supplementary material for the paper "Towards Visual Saliency Explanations of Face Verification". We demonstrate more results in this document to support the advantage and effectiveness of our proposed method. First, a visual comparison between CorrRISE and other explainable face verification methods in facial occlusion scenarios is presented. It further shows CorrRISE outperforms other approaches in localizing important regions. Then, more detailed quantitative experimental results are provided as a supplement to the information in the manuscript. Afterward, cross-model analysis has been performed. We show the visualization results after applying the CorrRISE method to four different face recognition models, which validates the generalization ability of the proposed explanation method as well as the effectiveness of our quantitative evaluation metrics.

A. Visual Comparison in Occlusion Scenarios

Fig. 1 compares the visualization results of our method with three classic and two state-of-the-art explanation methods under the facial occlusion scenarios. Similar regions have been highlighted by the produced saliency maps. It is shown that our proposed CorrRISE algorithm can most precisely highlight similar regions and exclude masked areas. Although xFace [3] generates compelling explanation results in standard verification scenarios, it often recognizes masked pixels as similar regions. MinPlus [4] obtains comparable results with ours, but this algorithm occasionally breaks down and fails to produce meaningful results in difficult verification cases, see row 4. It is also notable that CorrRISE generates much more accurate saliency maps than the straightforward adaptation of RISE.

B. Quantitative Experimental Results

The manuscript has shown comprehensive quantitative experimental results in the form of Tables on three differ-

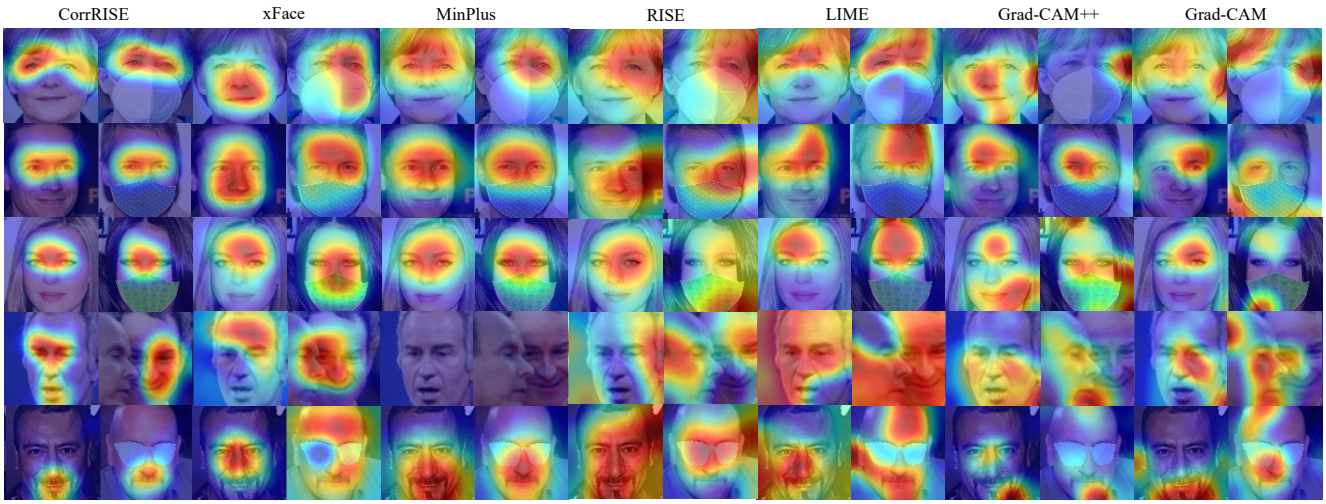

Figure 1. Visual results comparison among seven saliency map-based explanation methods in facial occlusion scenarios. CorrRISE provides more accurate saliency map explanations than current state-of-the-art methods. The importance increases from blue to red color.

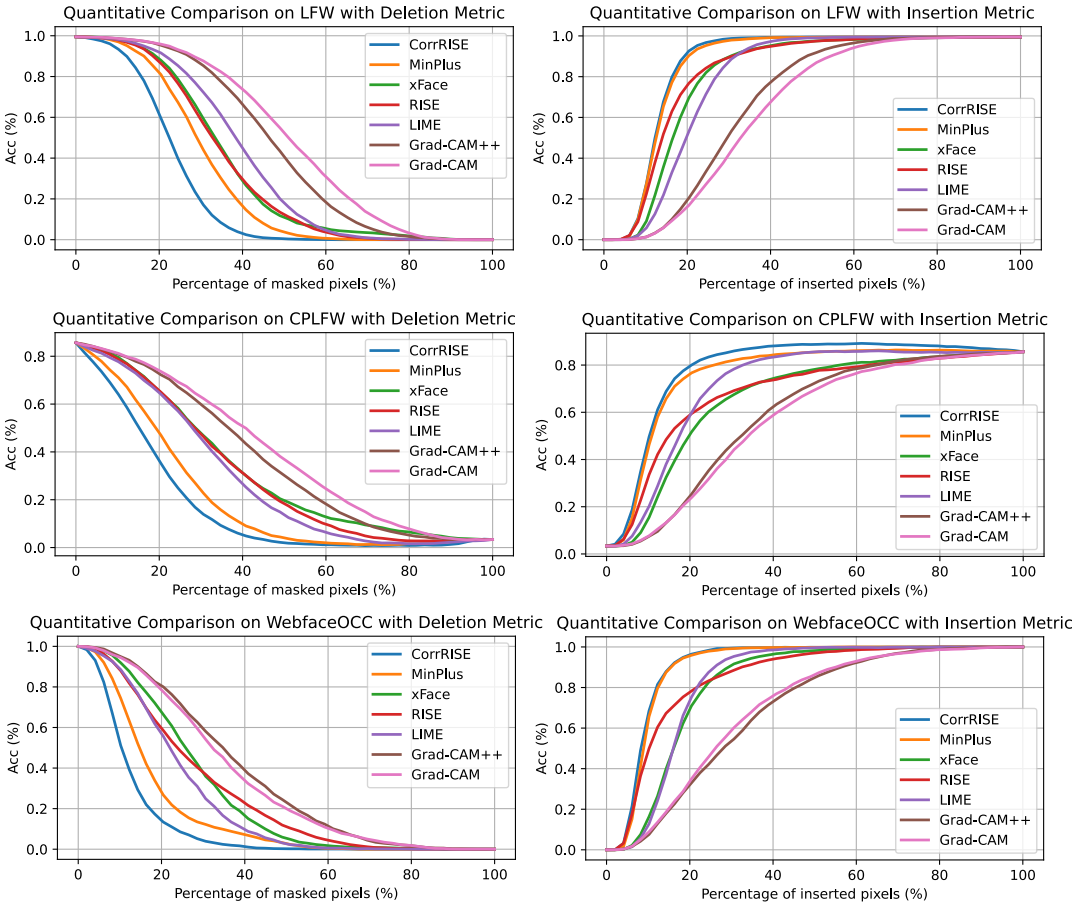

Figure 2. Quantitative comparison among the state-of-the-art XFR and XAI methods with Deletion and Insertion metrics. For the Deletion metric, the lower the better, and the opposite for the Insertion metric.

ent datasets, which represent three types of face verification scenarios. In this supplementary material, we demonstrate the comprehensive quantitative evaluation curves calculated with the Deletion and Insertion evaluation methodology. The values in the Table of the manuscript are the AUC scores of these curves.

Fig. 2 shows that the verification accuracy of the ArcFace model drops or rises the most rapidly on three datasets after masking or inserting sorted pixels following saliency maps generated by CorrRISE.

C. Visual Comparison across Face Recognition Models

In the manuscript, a cross-model quantitative comparison has been conducted using our proposed "Deletion" and "Insertion" metrics, which show that CorrRISE can produce accurate saliency maps regardless of which deep face recognition model to use. Fig. 3 further provides visualization samples of the saliency maps that CorrRISE created for different face recognition models. The saliency maps highlight similar regions, which validates that our proposed metrics are consistent with the visualization results and are reliable for a fair comparison among general saliency map-based explanation methods. It also shows that CorrRISE generalizes well across different face recognition models.

In Fig. 3, the produced saliency maps are very similar because all four FR models make correct predictions on the three examples. To further show the strong explainability of CorrRISE, Fig. 4 gives visual explanations for two opposite decisions made by ArcFace [1] and AdaFace [2] respectively. The former fails to recognize the given example and mistakenly verifies they are the same person, while the latter manages to give correct predictions. The saliency map produced by the CorrRISE method shows that the ArcFace model allocates high saliency values to similar regions between the non-matching pairs while indicating very low dissimilarity between them. On the contrary, the AdaFace model makes correct predictions because it believes there are strong dissimilar pixels between the given examples.

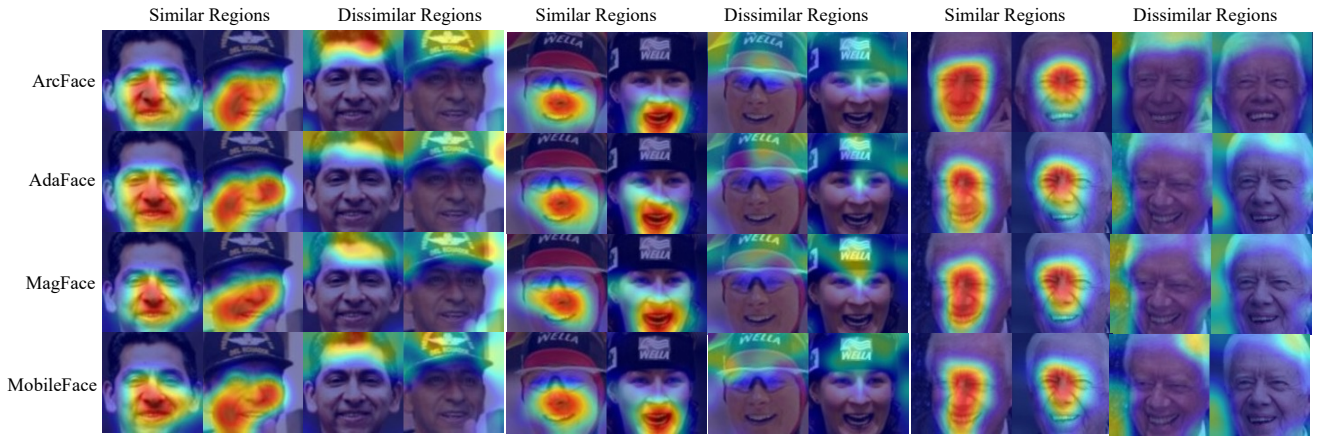

Figure 3. Visual saliency maps produced by CorrRISE for four different deep face recognition models. The visualization results are consistent with the quantitative metrics reported in the manuscript.

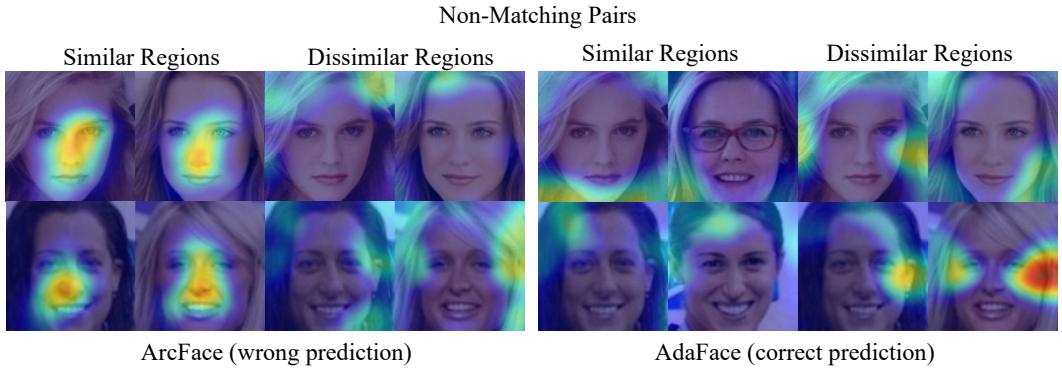

Figure 4. Visual explanation of two different decisions made by two face recognition models. ArcFace model mistakenly verifies the given example as matching while the AdaFace correctly recognizes they are non-matching.

This experiment is complementary to Fig. 3 and the reported sanity check in the manuscript, and further proves that CorrRISE is capable of providing meaningful explanations for different deep face-matching models.

References

[1] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019. 2

[2] Minchul Kim, Anil K Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18750–18759, 2022. 2

[3] Martin Knoche, Torben Teepe, Stefan Hörmann, and Gerhard Rigoll. Explainable model-agnostic similarity and confidence in face verification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 711–718, 2023. 1

[4] Domingo Mery. True black-box explanation in facial analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1596–1605, 2022. 1