Analysis of Bias-Mitigation Techniques and Evaluation Metrics in Face Recognition

Supervision: Yu Linghu, Prof. Dr. Manuel Günther

In face recognition systems, the task is to compare two facial images and determine whether they come from the same identity. The rapid advancement of deep learning has significantly enhanced the quality of face recognition systems. Face recognition via deep learning is always treated as a transfer learning task, where networks are trained to extract deep features from facial images, also known as face embeddings. Modern architectures and loss functions, such as ArcFace [Deng et al., 2022] and Adaface [Kim et al., 2022], have improved feature discriminability. Yet these systems often exhibit biases, particularly across different demographic groups. One possible reason is that training datasets like WebFace260M [Zhu et al., 2021] often lack demographic diversity, leading to imbalanced representation. Performance disparities across demographic groups have prompted significant research, with factors such as ethnicity [Vangara et al., 2019] and gender [Albiero et al., 2020] influencing accuracy. Various datasets such as Racial Faces in the Wild [Wang et al., 2019] and evolving fairness evaluation methods [Grother, 2022] highlight the complexity of achieving fair face recognition.

This thesis aims to analyze the current literature on bias mitigation techniques in face recognition, using multiple evaluation metrics to comprehensively assess their effectiveness. Bias mitigation methods include disentangling sensitive information from face embeddings via adversarial training [Gong et al., 2020, Morales et al., 2020] to learning less-biased face representations. Techniques like cluster-based large-margin local embedding loss [Huang et al., 2019], networks based on reinforcement learning [Wang and Deng, 2020], and other [Yang et al., 2021, Gong et al., 2021, Serna et al., 2022] show promise but often require fully labeled sensitive attributes. Other semi-supervised approaches [Qin, 2020] offer solutions for real-world data scenarios.

The primary objectives are to conduct a comprehensive review of current literature to provide an overview of existing bias mitigation techniques, and identify gaps for future research. A thorough literature review will focus on bias mitigation techniques, datasets, and fairness evaluation methods. A reasonable selection of such systems shall be implemented and evaluated using multiple evaluation metrics. Particularly, ROC curves, Fairness Discrepancy Rate [de Freitas Pereira and Marcel, 2021], the spread of False Match Rate, and False Non-Match Rate [Grother, 2022], shall be employed to assess the effectiveness of these techniques. Experiments will compare the performance of different methods, with networks trained using various approaches and evaluated across multiple demographics. The expected outcomes include a comprehensive understanding of current bias mitigation techniques, identification of the most effective techniques based on multiple evaluation metrics, and exploration of the possibility of extending them into multiple demographic scenarios.

Requirements

- A reasonable understanding of deep neural networks and their learning processes.
- Programming experience in Python and a deep learning framework, preferably PyTorch.
- Successfully passed the Deep Learning Course.
- Proficiency in written English.

References

[Albiero et al., 2020] Albiero, V., KS, K., Vangara, K., Zhang, K., King, M. C., and Bowyer, K. W. (2020). Analysis of gender inequality in face recognition accuracy. In *Winter Conference on Applications of Computer Vision (WACV) Workshops*, pages 81–89. IEEE/CVF.

[de Freitas Pereira and Marcel, 2021] de Freitas Pereira, T. and Marcel, S. (2021). Fairness in biometrics: A figure of merit to assess biometric verification systems. *Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, 4(1):19–29.

[Deng et al., 2022] Deng, J., Guo, J., Yang, J., Xue, N., Kotsia, I., and Zafeiriou, S. (2022). Arcface: Additive angular margin loss for deep face recognition. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 44(10):5962–5979.

[Gong et al., 2020] Gong, S., Liu, X., and Jain, A. K. (2020). Jointly de-biasing face recognition and demographic attribute estimation. In *European Conference on Computer Vision (ECCV)*. Springer.

[Gong et al., 2021] Gong, S., Liu, X., and Jain, A. K. (2021). Mitigating face recognition bias via group adaptive classifier. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Grother, 2022] Grother, P. (2022). Face recognition vendor test (FRVT) part 8: Summarizing demographic differentials. Technical report, National Institute of Standards and Technology (NIST).

[Huang et al., 2019] Huang, C., Li, Y., Loy, C. C., and Tang, X. (2019). Deep imbalanced learning for face recognition and attribute prediction. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*.

[Kim et al., 2022] Kim, M., Jain, A. K., and Liu, X. (2022). AdaFace: Quality adaptive margin for face recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Morales et al., 2020] Morales, A., Fierrez, J., Vera-Rodriguez, R., and Tolosana, R. (2020). SensitiveNets: Learning agnostic representations with application to face images. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*.

[Qin, 2020] Qin, H. (2020). Asymmetric rejection loss for fairer face recognition. *arXiv*, abs/2002.03276.

[Serna et al., 2022] Serna, I., Morales, A., Fierrez, J., and Obradovich, N. (2022). Sensitive loss: Improving accuracy and fairness of face representations with discrimination-aware deep learning. *Artificial Intelligence*.

[Vangara et al., 2019] Vangara, K., King, M. C., Albiero, V., Bowyer, K., et al. (2019). Characterizing the variability in face recognition accuracy relative to race. In *Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[Wang and Deng, 2020] Wang, M. and Deng, W. (2020). Mitigating bias in face recognition using skewness-aware reinforcement learning. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Wang et al., 2019] Wang, M., Deng, W., Hu, J., Tao, X., and Huang, Y. (2019). Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In *International Conference on Computer Vision (ICCV)*, pages 692–702. IEEE.

[Yang et al., 2021] Yang, Z., Zhu, X., Jiang, C., Liu, W., and Shen, L. (2021). RamFace: race adaptive margin based face recognition for racial bias mitigation. In *International Joint Conference on Biometrics (IJCB)*. IEEE.

[Zhu et al., 2021] Zhu, Z., Huang, G., Deng, J., Ye, Y., Huang, J., Chen, X., Zhu, J., Yang, T., Lu, J., Du, D., et al. (2021). WebFace260M: A benchmark unveiling the power of million-scale deep face recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.